

Multi-subject stochastic blockmodels with mixed effects for adaptive analysis of individual differences in human brain network cluster structure

Dragana M. Pavlović¹ | Bryan R.L. Guillaume |
Soroosh Afyouni | Thomas E. Nichols

Oxford Big Data Institute Li Ka Shing
Centre for Health Information and
Discovery Nuffield Department of
Population Health, University of Oxford,
Oxford, UK

Correspondence

Dragana M. Pavlović, Oxford Big Data
Institute Li Ka Shing Centre for Health
Information and Discovery Nuffield
Department of Population Health,
University of Oxford, Oxford, UK.
Email: dragana.pavlovic@bdi.ox.ac.uk

Funding information

Wellcome Trust, Grant/Award Number:
100309/Z/12/Z

Recently, there has been a renewed interest in the class of stochastic blockmodels (SBM) and their applications to multi-subject brain networks. In our most recent work, we have considered an extension of the classical SBM, termed heterogeneous SBM (Het-SBM), that models subject variability in the cluster-connectivity profiles through the addition of a logistic regression model with subject-specific covariates on the level of each block. Although this model has proved to be useful in both the clustering and inference aspects of multi-subject brain network data, including fleshing out differences in connectivity between patients and controls, it does not account for dependencies that may exist within subjects. To overcome this limitation, we propose an extension of Het-SBM, termed Het-Mixed-SBM, in which we model the within-subject dependencies by adding subject- and block-level random intercepts in the embedded logistic regression model. Using synthetic data, we investigate the accuracy of the partitions estimated by our proposed model as well as the validity of inference procedures based on the Wald and permutation tests. Finally, we illustrate the model by

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Statistica Neerlandica* published by John Wiley & Sons Ltd on behalf of VVS.

analyzing the resting-state fMRI networks of 99 healthy volunteers from the Human Connectome Project (HCP) using covariates like age, gender, and IQ to explain the clustering patterns observed in the data.

KEYWORDS

community detection, integrated classification likelihood criterion, mixed effects, mixture models, multi-subject network analysis, network inference, permutation test, stochastic block model, stochastic blockmodel, variational, Wald test

1 | INTRODUCTION

Network analysis and clustering methods have emerged as powerful platforms for investigating how some units of interest (i.e., nodes) mutually interact to facilitate the functioning of a system as a whole. As illustrated in Figure 1, data from resting-state fMRI can be abstracted to network like objects so that the anatomically defined regions of interest (ROIs) represent n number of nodes and the binary-valued links between them represent edges (with $\{1, 0\}$ denoting the presence and absence of an edge, respectively).

The underlying topology of a brain network, which we refer to as the cluster structure, is revealed through its decomposition into Q number of clusters, which flesh out the most prominent connectivity profiles. Generally, this topology is estimated using one of the many existing clustering methods, the majority of which are based on the optimization of an objective function called modularity (e.g., the Newman spectral and fast Louvain algorithms; Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006; Bassett & Bullmore, 2009; Bullmore & Sporns, 2009; Hilgetag, Burns, O'Neill, Scannell, & Young, 2000; Pan, Chatterjee, & Sinha, 2010; Rubinov & Sporns, 2010; Sporns, Chialvo, Kaiser, & Hilgetag, 2004). Nevertheless, one of the limitations of methods based on modularity is that they are confined to the identification of a limited range of patterns of connectivity, which typically fall into the framework of the “modular organization” (in which, the within-cluster connectivities are much higher than the between-cluster connectivities). In an extensive set of simulations (Pavlovic et al., 2020), we have benchmarked these methods on modular topologies but also on topologies with a core-modular structure, which diverge from a modular architecture by combining a subset of clusters with a modular like organization and a subset of strongly interconnected clusters. On these simulations, the modular algorithms showed a very high degree of inaccuracy in retrieving the correct cluster labels when the structure was heterogeneous modular but especially when the topology was nonmodular. A more accurate alternative to these methods is contained to the class of probabilistic network clustering methods called stochastic blockmodel (SBM), which were developed in the work of Snijders and Nowicki (Nowicki & Snijders, 2001; Snijders & Nowicki, 1997). The SBMs use a framework of mixture models to explain variations in the distribution of edges on a single network. Depending on the edge values, one would select a unique family of probability density (e.g., the Bernoulli density), which would parametrize the connection probabilities within each cluster and between each cluster pair. In the Bernoulli density example, these parameters represent the mean values in the cluster structure π , which are given as a $Q \times Q$ matrix of probabilities. The participation of each Bernoulli cluster in the overall mixture of densities is controlled by the parameter α , which

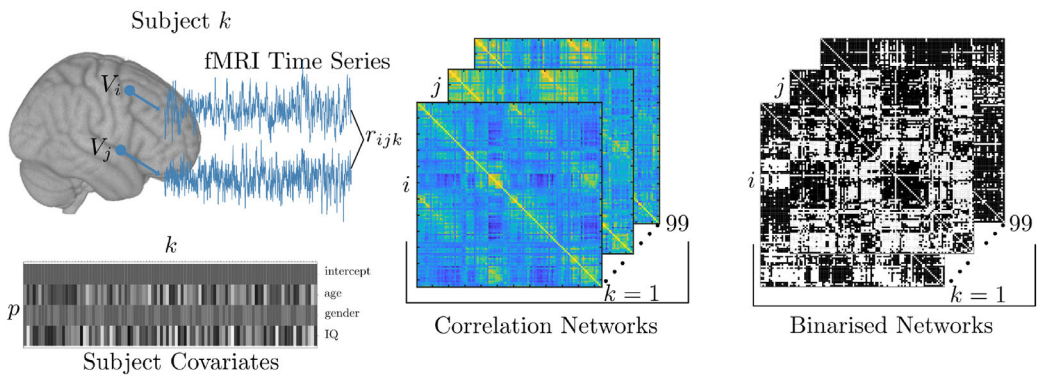


FIGURE 1 Illustration of the main steps involved in the derivation of multi-subject binary networks from resting-state fMRI data. The nodes, labeled as V_i and V_j , represent n regions of interest usually based on some predefined brain parcellations or anatomical atlases. The time-series data in each such region is represented by the mean time series across its constituent voxels. There are typically four ways to generate binary networks from these time series. (i) The first way is to compute Pearson's correlations between the nodal time courses, then transform them using Fisher's z scores and finally binarize them based on the series of tests $\{H_0: r_{ijk} = 0; H_1: r_{ijk} > 0\}$ in which the significant edges take value one and the nonsignificant edges take value zero. To carry out these tests, we have used the “xDF” method recently proposed by Afyouni, Smith, and Nichols (2019), which accounts for the autocorrelation in regional time series, including the instantaneous and lagged cross-correlations, and computes more accurate variances for each edge. (ii) Another way is to directly threshold the correlation matrices according to some predefined correlation value (e.g., 0.8), which treats only the strongest correlations as edges. (iii) The third way is to decompose the nodal time series into discrete wavelets and compute the correlations between the wavelet coefficients within a low-frequency scale for which fMRI dynamics of neuronal origin are the most prominent (e.g., < 0.1 Hz). (iv) The fourth way is to use partial correlation, which takes each nodal time series pair and regresses out the other $n - 2$ nodal time courses from them. At the end of any of the four pipelines, each subject is awarded a binary symmetric adjacency matrix \mathbf{x}_k , and subject-specific covariates are collected alongside the resting-state fMRI data

is a vector of length Q . By parameterizing each element in a cluster structure, the SBM model captures a wide range of topologies, and it is not limited to the stringent assumptions of modular organizations. Our SBM application on the *C. elegans* network (Pavlovic, Vértes, Bullmore, Schafer, & Nichols, 2014) captured its locomotion circuit and delineated clusters of sensory neurons maintaining the forward and backward locomotion, suggesting that the model can be used to identify clusters with functionally similar nodes.

To obtain the estimates of $\boldsymbol{\pi}$, $\boldsymbol{\alpha}$, and the latent cluster labels \mathbf{z} , Snijders and Nowicki (1997) considered optimizations based on the Gibbs sampling and expectation maximization (EM). More recently, Daudin, Picard, and Robin (2008) introduced a frequentist variational approximation as optimization strategy and derived a criterion (Biernacki, Celeux, & Govaert, 2000) which acts as a goodness of fit measure for each network partition. Since these developments, the SBMs have been adapted for various biological data sets, including the overlapping SBM (Latouche, Birmelé, & Ambroise, 2011; Latouche, Birmelé, & Ambroise, 2014), the closely related mixed-membership SBM (Airoldi, Blei, Fienberg, & Xing, 2008), the bayesian SBM (Côme & Latouche, 2015; Latouche, Birmelé, & Ambroise, 2012), the online SBM (Zanghi, Ambroise, & Miele, 2008), the SBM with nodal features (Zanghi, Volant, & Ambroise, 2010), and the SBM with edge features by Mariadassou, Robin, and Vacher (2010). SBMs have been applied in the analysis of macaque anatomical cortex (Picard, Miele, Daudin, Cottret, & Robin, 2009), *Caenorhabditis elegans* brain network (Pavlovic et al., 2014) and human brain networks (Hinne, Ekman, Janssen,

Heskes, & van Gerven, 2015; Moyer et al., 2015; Pavlovic, 2015). Of note, SBM hybrids have also been considered in the analysis of dynamic connectivity (Matias & Miele, 2017) and in neuroimaging (Robinson, Atlas, & Wager, 2015). Theoretical properties of the SBMs have additionally been studied in the work of Bickel, Choi, Chang, and Zhang (2013), Choi, Wolfe, and Airolidi (2012), Ambroise and Matias (2012), Wolfe and Olhede (2013), Olhede and Wolfe (2014), and Gao, Lu, and Zhou (2015) to mention a few.

Among this literature, we especially highlight the work of Mariadassou et al. (2010) who developed the generalized SBMs for the analysis of a single network. This work extends the variational fitting procedure of SBMs for the exponential family of densities and uses generalized linear models with edge-based covariate values to explain the network cluster structure π . The interplay between the cluster structure and covariates can be such that the effects of covariates are the same across the entire cluster structure (i.e., homogeneous effects) or affect each element of the cluster structure with different intensity (i.e., heterogeneous effects). Building upon the work of Mariadassou et al. (2010), we have proposed three multi-subject SBMs (MS-SBMs) among which we especially highlight the heterogeneous-effects SBM (Het-SBM; Pavlovic et al., 2020), which serves as a basis for the development of the new model in this work.

Het-SBM assumes that the cluster labels are the same for all subjects but uses subject-specific covariates like age or gender in a logistic regression model to explain between-subject variations in the cluster structure. To intuitively understand this model, in Figure 2, we showcase a straightforward example in which the age effect of three subjects acts differently on the connectivity in each block. For example, there is a decrease of connectivity in Block (1, 1) with increasing age, and the opposite directional effect is observed in Block (1, 2). By tying each covariate effect to a block cell, the model can capture a wide range of cluster structures. For example, the first subject has a modular topology as the within-block connections are higher than those between-block connections. By contrast, the third subject has a topology that resembles a disassortative mixing as the bulk of connections in the profile of Block 1 goes to maintaining ties with Blocks 2 and 3. The estimation of model parameters related to the logistic regression is based on Firth-type estimators (Firth, 1993) which belong to a class of bias preventative procedures. At first glance, the assumption of common cluster labels across subjects may appear too strong and potentially discount the use of covariates in the estimation of cluster labels. However, since the effects of the covariates are tied to each element of the cluster structure, it is possible to encounter examples in which the effects of covariates iron out the differences in the connectivity profiles across the clusters and this information can be used toward a more accurate estimation of cluster labels. Thus, it is not advisable to separate the cluster label estimation from the logistic regression. In addition to its clustering abilities, the Het-SBM model also provides inference tools to detect group differences in the connectivity rates or, more generally, effects of covariates.

However, although the Het-SBM model is useful for the analysis of independent multi-subject networks, it may not be appropriate in cases where some form of dependence exists in the data. This may be generally attributed to two sources. First, if the covariates do not fully explain the intersubject differences in connectivity rates, there may still be some dependence within the individual elements of a block structure. For example, the prevalence of edges in one subject's block (e.g., the block (q, l)) may be consistently overestimated and, therefore, this lack of fit may be randomly distributed over subjects. Second, the data may comprise more than one network per each subject (i.e., visits) inducing repeated measures correlations. The latter type of data typically occurs in studies with multiple visits of subjects, who are either scanned at different time points or after being exposed to different experimental conditions, such as different drug effects (e.g., a crossover design). Other examples of such a type of data may include the combined outcomes of

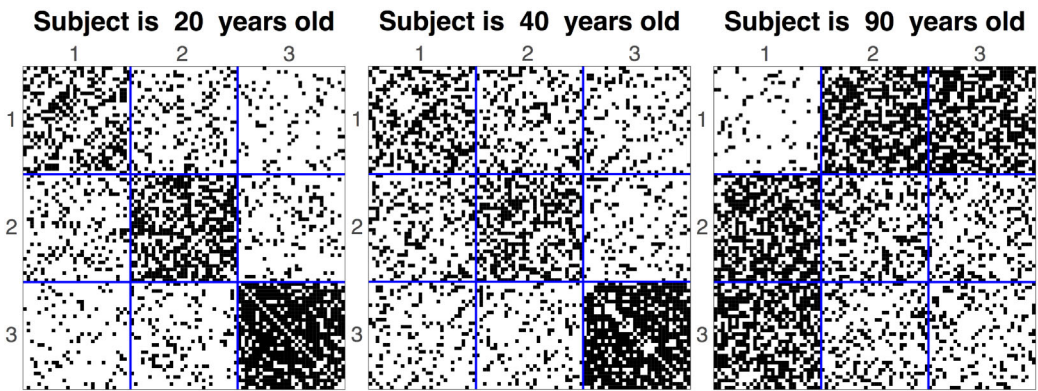


FIGURE 2 Illustration of heterogeneous SBM (Het-SBM). Simulated data are available for three subjects aged 20, 40, and 90 years, and each subject's network is represented as a reorganized adjacency matrix comprised of three blocks, labeled numerically (1–3). In Het-SBM, the subject-level covariates can affect the connectivity rates specifically within each block or between each pair of blocks. The effect of age is seen as a locally heterogeneous increase or decrease of connectivity in each block. For example, the intrablock connectivity in Block (1, 1) decreases as a function of age, whereas the interblock connectivity in Block (1, 2) increases with age

different imaging modalities or data that combine different classes of connectivity (e.g., structural and functional), where the goal would be to answer questions regarding the similarities of their networks' organizations.

In the present work, we continue the characterization of the human brain network topology by extending previous works in several ways. First, we derive the Het-Mixed-SBM model (see Pavlovic, 2015, chapter 5), which uses the heterogeneous fixed effects from the Het-SBM model and poses block- and subject-level random effects. Second, we outline its estimation strategy and consider the parametric Wald test and permutation test based on the Wald statistics. Third, we validate the model on a range of simulation scenarios where we benchmark this model against Het-SBM and showcase instances where Het-Mixed-SBM can provide more accurate clustering. In another set of simulations, we show the control of false positives and, finally, we consider an application to a resting-state fMRI study obtained from the Human Connectome Project (Van Essen et al., 2012) with 99 healthy subjects and 114 nodes. This data set has only one visit per each subject and, among the wealth of information available for these subjects, we focus on covariates such as age, gender, and IQ score, which are often linked to between-subject variations.

2 | METHOD

Hereafter, we will be using the classical convention of roman capital letters to denote random variables and lower case letters to denote their observed realizations. Scalar and nonscalar values will be denoted by light and boldface fonts, respectively.

2.1 | Heterogeneous mixed-effects stochastic blockmodel (Het-Mixed-SBM)

Let the set of nodes, labeled as $\{V_1, \dots, V_n\}$, be divided into Q unknown (latent) blocks or clusters. For a fixed number of blocks, Q , the indices $q, l \in \{1, \dots, Q\}$ are used to label the individual

blocks. The block membership of a particular node V_i will be indicated by a random vector $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iQ})$, whose elements Z_{iq} take the value 1 if $V_i \in q$ th group and 0 otherwise. For example, fixing $Q = 3$, the estimated $\hat{\mathbf{z}}_i = (0, 1, 0)$ indicates that the node V_i is in the second cluster. Pooling this information across nodes, the $n \times Q$ random matrix \mathbf{Z} can be defined such that the vectors \mathbf{Z}_i are mutually independent and follow a categorical density with Q possible outcomes

$$\mathbf{Z}_i \sim \text{Categorical}(Q, \boldsymbol{\alpha}), \quad (1)$$

where $\boldsymbol{\alpha}$ is a $1 \times Q$ dimensional vector of success probabilities $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q)$ with a specific α_q being the probability that a randomly selected node falls into the q th block, and $\sum_{q=1}^Q \alpha_q = 1$. Here, it is important to note that the choice of categorical (or equivalently single-trial multinomial) density implies that the fitted blocks form a partition of all nodes, in which, each node belongs to only one block (i.e., disjoint blocks). This is formally noted as $\sum_{q=1}^Q z_{iq} = 1$. Writing $\mathbf{Z} = ((Z_{iq}))_{1 \leq i \leq n, 1 \leq q \leq Q}$ for the $n \times Q$ matrix, the probability mass function is given as

$$f(\mathbf{z}; \boldsymbol{\alpha}) = \prod_{q=1}^Q \prod_{i=1}^n \alpha_q^{z_{iq}}. \quad (2)$$

The random variable X_{ijkt} represents the possibility of an edge between the nodes V_i and V_j in the k th subject and its t th measurement. Hence, x_{ijkt} denotes a binary realization of X_{ijkt} with 1 being an edge and 0 no edge. Therefore, for the k th subject and t th measurement, $\mathbf{X}_{kt} = ((X_{ijkt}))_{1 \leq i \neq j \leq n}$ denotes an $n \times n$ random, symmetric adjacency matrix with elements X_{ijkt} , and, for a total of K subjects and T measurements, \mathbf{X} denotes the set of random matrices $\mathbf{X} = \{\mathbf{X}_{kt} : k \in \{1, \dots, K\}, t \in \{1, \dots, T\}\}$. The individual matrices $\mathbf{x}_{kt} = ((x_{ijkt}))_{1 \leq i \neq j \leq n}$ are assumed to be undirected, without self-connected nodes and without multiple edges between the nodes. Hence, they are binary and symmetric matrices with 0s on their principal diagonal and a total of $n(n-1)/2$ data points. Conditional on the cluster labels, the edges follow a Bernoulli density

$$X_{ijkt} | Z_{iq} = 1, Z_{jl} = 1, R_{qlk} = r_{qlk} \sim \text{Bernoulli}(\pi_{qlkt}), \quad (3)$$

$$\log \left(\frac{\pi_{qlkt}}{1 - \pi_{qlkt}} \right) = \mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql} + r_{qlk}, \quad \pi_{qlkt} = \frac{e^{\mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql} + r_{qlk}}}{1 + e^{\mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql} + r_{qlk}}}, \quad (4)$$

$$R_{qlk} \stackrel{iid}{\sim} \text{Normal}(0, \sigma_{ql}^2), \quad (5)$$

where R_{qlk} represents a subject-specific random intercept for the block (q, l) , \mathbf{d}_{kt} is a $P \times 1$ vector of covariates associated with the k th subject and its t th measurement, and $\boldsymbol{\beta}_{ql}$ is a $P \times 1$ vector of regression coefficients for the block (q, l) . The total of $Q(Q+1)/2$ individual regression vectors $\boldsymbol{\beta}_{ql}$ is collectively noted as $\boldsymbol{\beta}$. We note that σ_{ql}^2 is the random effect variance of each block or block-to-block regression and it is collectively noted as $\boldsymbol{\sigma}^2 = ((\sigma_{ql}^2))_{1 \leq q, l \leq Q}$.

In particular, the probability mass function of \mathbf{x} given \mathbf{z} can be written as

$$f(\mathbf{x} | \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \prod_{k=1}^K \prod_{q=1}^Q \prod_{l=1}^Q \int_{-\infty}^{+\infty} \left(\left[\sum_{i=1}^n \sum_{j>i}^n \sum_{t=1}^T [f(X_{ijkt} | z_{iq}, z_{jl}, r_{qlk}; \boldsymbol{\beta}_{ql})]^{z_{iq} z_{jl}} \right] \times f(r_{qlk}; \sigma_{ql}^2) \right) dr_{qlk}. \quad (6)$$

2.1.1 | Estimation

The model parameters are estimated by optimizing the variational bound $\mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ defined as

$$\mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \mathbb{E}_{f^*} [\log f(\mathbf{x}|\mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)] + \mathbb{E}_{f^*} [\log f(\mathbf{Z}; \boldsymbol{\alpha})] - \mathbb{E}_{f^*} [\log f^*(\mathbf{Z}; \boldsymbol{\tau})], \quad (7)$$

where $f^*(\mathbf{z}; \boldsymbol{\tau})$ is the variational density with parameter $\boldsymbol{\tau}$ ($n \times Q$ matrix of cluster probabilities), and it denotes the parametric family that is the closest (in the Kullback–Leibler sense) to $f(\mathbf{z}|\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha})$. The evaluation of Equation (7) requires taking the expectation of \mathbf{Z} with respect to its variational density $f^*(\mathbf{z}; \boldsymbol{\tau})$, which is taken to be a product of the individual densities of \mathbf{Z}_i . Each density is categorical with block-specific probabilities that are independent in each node,

$$f^*(\mathbf{z}; \boldsymbol{\tau}) = \prod_{i=1}^n \prod_{q=1}^Q \tau_{iq}^{z_{iq}}, \quad (8)$$

where $\sum_{q=1}^Q \tau_{iq} = 1$, $\mathbb{E}_{f^*}(Z_{iq}) = \tau_{iq}$ and $\mathbb{E}_{f^*}(Z_{iq}Z_{jl}) = \tau_{iq}\tau_{jl}$. In particular, τ_{iq} is the strength of evidence that a node V_i is a member of block q having observed the data. Let us suppose an example in which the node V_i has a vector $\hat{\boldsymbol{\tau}}_i = (0.1, 0.01, 0.89)$, expressing the likelihood of its membership in three clusters. Since this node has the strongest affiliation to the third block, its $\hat{\mathbf{z}}_i = (0, 0, 1)$. For simplicity of notation, we take advantage of the fact that $\boldsymbol{\beta}_{ql} = \boldsymbol{\beta}_{lq}$ and define γ_{ijql} as

$$\gamma_{ijql} = \begin{cases} \tau_{iq}\tau_{jq} & \text{if } q = l \\ \tau_{iq}\tau_{jl} + \tau_{il}\tau_{jq} & \text{if } q < l. \end{cases} \quad (9)$$

According to this, the variational bound in Equation (7) is given as

$$\begin{aligned} \mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) &= \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \sum_{q=1}^Q \sum_{l \geq q}^Q \gamma_{ijql} \mathbf{x}_{ijkt} \mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql} \\ &\quad - \frac{K}{2} \sum_{q=1}^Q \sum_{l \geq q}^Q \log(2\Pi\sigma_{ql}^2) \\ &\quad + \sum_{k=1}^K \sum_{q=1}^Q \sum_{l \geq q}^Q \log \int_{-\infty}^{+\infty} e^{h_{qlk}} dr_{qlk} \\ &\quad + \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log(\alpha_q) - \sum_{i=1}^n \sum_{q=1}^Q \tau_{iq} \log(\tau_{iq}), \end{aligned} \quad (10)$$

$$h_{qlk} = \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \mathbf{x}_{ijkt} r_{qlk} + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \log(1 - \pi_{qlkt}) - \frac{r_{qlk}^2}{2\sigma_{ql}^2}. \quad (11)$$

For a fixed Q , we want to maximize the variational bound defined by Equation (10) with respect to the variational parameter τ as well as the parameters α , β , and σ^2 . The optimal variational parameter τ satisfies the fixed point relations

$$\hat{\tau}_{iq} \propto \hat{\alpha}_q \left(\prod_{k=1}^K \prod_{l=1}^Q \exp \left(\frac{\hat{I}_{iqlk}}{\hat{I}_{qlk1}} \right) \right) \prod_{k=1}^K \prod_{j \neq i}^n \prod_{t=1}^T \prod_{l=1}^Q \exp \left(\hat{\tau}_{jl} x_{ijkt} \mathbf{d}_{kt}^\top \hat{\beta}_{ql} \right), \quad (12)$$

where we use the notation I_{qlk1} and I_{iqlk} to denote the following integral expressions

$$\begin{aligned} \hat{I}_{qlk1} &= \int_{-\infty}^{+\infty} e^{\hat{h}_{qlk}} dr_{qlk}, \\ \hat{I}_{iqlk} &= \frac{\partial \hat{I}_{qlk1}}{\partial \tau_{iq}} = \sum_{j \neq i}^n \hat{\tau}_{jl} \sum_{t=1}^T x_{ijkt} \int_{-\infty}^{+\infty} e^{\hat{h}_{qlk}} r_{qlk} dr_{qlk} \\ &\quad + \sum_{i \neq j}^n \hat{\tau}_{jl} \sum_{t=1}^T \int_{-\infty}^{+\infty} e^{\hat{h}_{qlk}} \log(1 - \hat{\pi}_{qlkt}) dr_{qlk}. \end{aligned} \quad (13)$$

The estimate of each element of α is given as

$$\hat{\alpha}_q = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_{iq}. \quad (14)$$

We next turn our attention to the optimization of the variational bound for β and σ^2 . The estimates of β_{ql} and σ_{ql}^2 can be found with the Newton–Raphson formula so that, for the (m) th step, we have

$$\begin{pmatrix} \beta_{ql} \\ \sigma_{ql}^2 \end{pmatrix}^{(m)} = \begin{pmatrix} \beta_{ql} \\ \sigma_{ql}^2 \end{pmatrix}^{(m-1)} + \mathcal{J}(\beta_{ql}^{(m-1)}, \sigma_{ql}^{2(m-1)})^{-1} \mathbf{U}(\beta_{ql}^{(m-1)}, \sigma_{ql}^{2(m-1)}). \quad (15)$$

The derivation details of the Fisher information matrix and the score can be found in Appendix A1.

For some initial values for τ^0 , we iteratively update the model parameters according to the two steps procedure

1. $(\alpha^{(h+1)}, \beta^{(h+1)}, \sigma^{2(h+1)}) = \underset{(\alpha, \beta, \sigma^2)}{\operatorname{argmax}} [\mathcal{J}(f^*(\mathbf{z}; \tau^{(h)}); \alpha, \beta, \sigma^2)],$
2. $\tau^{(h+1)} = \underset{\tau}{\operatorname{argmax}} [\mathcal{J}(f^*(\mathbf{z}; \tau); \alpha^{(h+1)}, \beta^{(h+1)}, \sigma^{2(h+1)})].$

In the first step, α is updated according to Equation (14) while β and σ^2 are updated according to the Newton–Raphson method outlined in Equation (15). In the second step, τ is updated according to Equation (12). The algorithm cycles through these two steps until convergence is obtained. The convergence is measured by the relative changes of the parameter estimates and the improvement of the variational bound. The Newton–Raphson algorithm uses a naïve estimate of $\hat{\pi}_{ql}$ as starting values for the intercept β_{ql}^{INT} (i.e., $\hat{\beta}_{ql}^{\text{INT}} = \log(\hat{\pi}_{ql}/(1 - \hat{\pi}_{ql}))$). In particular, for a given block element (i.e., Block (q, l)), $\hat{\pi}_{ql}$ is the ratio of the sum of its observed edges across the K subjects and T measurements and the total number of possible edges. The starting values for σ^2 are based on a strategy outlined in Demidenko (2004) whose objective is to estimate the random effects r_{qlk} and their sample variance, from which, we finally obtain the initial estimate

of σ^2 . The full details of this procedure can be found in Appendix B1 and further implementation details can be found in Appendix C1.

2.1.2 | Model selection

The integrated classification likelihood (ICL) criterion (Biernacki et al., 1998) is one of the most common procedures for determining the optimal number of clusters in mixture models, where the main goal is to cluster network data (Daudin et al., 2008; Mariadassou et al., 2010; Pavlovic et al., 2014). The benefit of this approach is that it favors the models with well-separated clusters, and it penalizes for model complexity, which ensures parsimony. Using m_Q to denote a candidate model with Q clusters, we derive the ICL criterion from $\log f(\mathbf{x}, \mathbf{z} | m_Q)$, which can be viewed as a sum of $\log f(\mathbf{x} | \mathbf{z}, m_Q)$ and $\log f(\mathbf{z} | m_Q)$. The individual components $\log f(\mathbf{x} | \mathbf{z}, m_Q)$ and $\log f(\mathbf{z} | m_Q)$ are each approximated with the Schwarz criterion (Schwarz, 1978). In the first component, $\log f(\mathbf{x} | \mathbf{z}, m_Q)$, the total number of parameters is $\frac{Q(Q+1)}{2}P$ in $\boldsymbol{\beta}$ and $\frac{Q(Q+1)}{2}$ in σ^2 and the total number of data points in \mathbf{x} is $\frac{n(n-1)}{2}KT$. In the second component, $\log f(\mathbf{z} | m_Q)$, the total number of parameters in $\boldsymbol{\alpha}$ is $Q - 1$, while the total number of data points is n . With this, the ICL criterion of Het-Mixed-SBM is defined as

$$\begin{aligned} \text{ICL}(m_Q) = & \log f(\mathbf{x}, \hat{\mathbf{z}} | m_Q, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2) \\ & - \frac{1}{2} \left(\frac{Q(Q+1)}{2} (P+1) \right) \log \left[\frac{n(n-1)}{2} KT \right] \\ & - \frac{Q-1}{2} \log[n]. \end{aligned} \quad (16)$$

2.1.3 | Inference

Het-Mixed SBM offers the possibility to estimate a common cluster structure across subjects that can serve as a common ground for making comparisons between them. Since it poses a mixed effect logistic regression model on each element of the block structure, its methodological framework can be used to estimate differences between groups of subjects or effects of covariates on the connectivity rate at each block. More precisely, we can statistically test if these quantities are different from 0 or, in a more general sense, if linear combinations of these quantities are different from a specified constant. In the continuation of this section, we consider the Wald test to make such a kind of hypothesis testing and assume it to be conditional on $\hat{\mathbf{z}}$.

Wald test.

The null hypothesis is given as $\mathcal{H}_0 : \mathbf{L}_{ql} \boldsymbol{\beta}_{ql} = \mathbf{b}_{ql0}$ and the Wald statistic can then take the form

$$W_{ql} = (\mathbf{L}_{ql} \hat{\boldsymbol{\beta}}_{ql} - \mathbf{b}_{ql0})^\top (\mathbf{L}_{ql} \boldsymbol{\mathcal{I}}_1^{-1} (\hat{\boldsymbol{\beta}}_{ql}) \mathbf{L}_{ql}^\top)^{-1} (\mathbf{L}_{ql} \hat{\boldsymbol{\beta}}_{ql} - \mathbf{b}_{ql0}) / c_{ql}, \quad (17)$$

where \mathbf{L}_{ql} is a matrix (or a vector) defining the combination of the parameters (contrast) tested and c_{ql} is the rank of \mathbf{L}_{ql} . Asymptotically, W_{ql} follows a $\chi^2_{c_{ql}}$ distribution. If \mathbf{L}_{ql} is a vector, then

$$W_{ql}^* = \frac{\mathbf{L}_{ql} \hat{\boldsymbol{\beta}}_{ql} - \mathbf{b}_{ql0}}{\sqrt{\mathbf{L}_{ql} \boldsymbol{\mathcal{I}}_1^{-1} (\hat{\boldsymbol{\beta}}_{ql}) \mathbf{L}_{ql}^\top}}, \quad (18)$$

asymptotically follows a standard normal distribution. The standard errors of the model parameters are estimated with the Fisher information matrix (see Appendix A1, Equation (A9)).

Multiple testing.

Inference procedures for the block level parameters comprise a multiple testing problem as we are effectively making $Q(Q+1)/2$ individual tests. To control the familywise error rate (FWE), defined as the probability of making at least one Type I error, we use the Bonferroni correction (Holm, 1979). This correction is valid for any dependence structure and is easy to apply: instead of using a nominal α_0 significance value (e.g., 0.05), α_0/n_T is used instead, where n_T is the number of tests (here, $n_T = Q(Q+1)/2$).

Permutation testing.

In addition to the Wald test, which depends on asymptotic sampling distributions, permutation tests are also considered (Good, 2000). Permutation tests are based on the premise that, under the null hypothesis, the data can be exchanged without altering its distribution. This implies that the null distribution of any test statistic can be found empirically through repeated evaluations of randomly rearranged (or permuted) data. In the context of Het-Mixed-SBM, we use the permutation strategy proposed by Potter (2005). In this approach, the covariate of interest is regressed on the remaining nuisance covariates (using a linear regression model), and the residuals from this model are used in place of the original covariate. These residuals are then permuted M times and, for each permutation $b = 1, \dots, M$, the logistic mixed-effect model is refitted yielding a permuted Wald test statistic w_{qlb} . The p -value of the original Wald test statistic w_{ql0} (recomputed by refitting the logistic mixed-effect model with the residual covariate) is then computed as

$$P(W_{ql0} \geq w_{ql0} | H_0) = \frac{\sum_{b=0}^M I(w_{qlb} \geq w_{ql0})}{M+1}, \quad (19)$$

where $I(\cdot)$ is the indicator function.

Improved multiple testing procedures with permutation.

The Bonferroni correction for controlling the FWE is known to be conservative in the presence of dependence across tests. Conveniently, the use of permutation can provide a less conservative control of the FWE by comparing the original statistics to the null distribution of the maximum statistic across blocks (Westfall & Young, 1993). For the original Wald statistic w_{ql0} , the FWE corrected p -value based on this procedure is given as

$$P(\max [(W_{ql})_{1 \leq q, l \leq Q}] \geq w_{ql0} | H_0) = \frac{\sum_{b=0}^M I(\max [(w_{qlb})_{1 \leq q, l \leq Q}] \geq w_{ql0})}{M+1}. \quad (20)$$

2.2 | Heterogeneous SBM (Het-SBM)

Het-SBM can be seen as Het-Mixed-SBM without random effects. Formally, for this model, we have

$$X_{ijk} | Z_{iq} = 1, Z_{jl} = 1 \sim \text{Bernoulli}(\pi_{qlk})$$

$$\log \left(\frac{\pi_{qlk}}{1 - \pi_{qlk}} \right) = \mathbf{d}_k^T \boldsymbol{\beta}_{ql} \quad \text{or equivalently} \quad \pi_{qlk} = \frac{1}{1 + e^{-\mathbf{d}_k^T \boldsymbol{\beta}_{ql}}}, \quad (21)$$

where π_{qlk} is the connectivity of block (q, l) associated with subject k , \mathbf{d}_k^\top is the $1 \times P$ vector of covariates associated with subject k (typically the first element will be 1, representing the intercept), and β_{ql} is a $P \times 1$ vector of regression parameters. Its estimation details and inference strategies can be found in Pavlovic et al. (2020).

3 | SIMULATION STUDY

This section first details the methodology of each simulation setting, labeled *Simulation I* and *Simulation II*, and then it proceeds to describe their results. Given that there is an overlap in the methodology of *Simulation I* and *Simulation II*, they are discussed jointly (see Section 3.1), while their results are treated separately (see Sections 3.2 and 3.3).

3.1 | Methodology of Simulations I and II

Figure 3 illustrates a bird's eye view of each simulation pipeline. Each simulation pipeline is broken down into four components: *goals*, *input parameters*, *models*, and *evaluation strategies*.

3.1.1 | Goals.

The goal of *Simulation I* is to investigate the accuracy of Het-SBM and Het-Mixed-SBM in estimating the correct cluster labels under the presence of random effects in the data, and to gauge how the random effects can influence this accuracy. The goal of *Simulation II* is to investigate, for both Het-SBM and Het-Mixed-SBM, the accuracy of the Wald parametric test and the nonparametric test (permutation test based on the Wald statistics) to control the Type I error rate and to give some recommendations for real data applications.

3.1.2 | Input parameters.

The input parameters in both simulations are the cluster structure, the fixed-effects sizes, the random-effects variances, the proportion design (i.e., the cluster sizes), the network size, the number of subjects, and the number of visits. In *Simulation I*, the cluster structures considered are the heterogeneous modular (Het-Modular), core-modular, and weak structures shown in Figure 3. The heterogeneous modular structure is similar to the commonly assumed perfectly modular structure in the sense that the within-cluster connectivities are larger than the between-cluster connectivities. However, it allows the connectivities to vary within and between clusters while the perfectly modular structure imposes them to be constant. The core-modular structure is a hybrid structure where some parts of the network are modular, while the remaining parts contain clusters with high mutual connectivity patterns that comprise, therefore, a densely integrated core of clusters. For example, Blocks 1 and 2 are core clusters because they are densely connected, while Blocks 2 and 3 are modular because their within-cluster connectivities are much higher than their between-cluster connectivities. These two structures were chosen because of the extensive evidence that brain networks demonstrate at least some degree of modularity (Meunier,

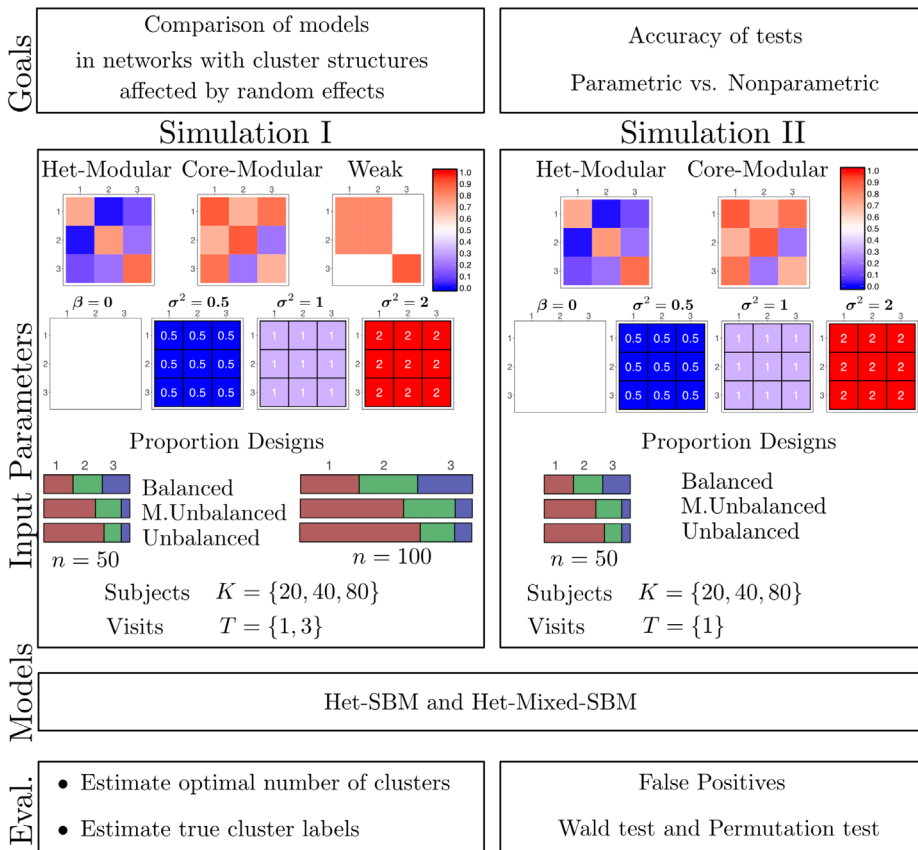


FIGURE 3 Simulation pipelines: goals, input parameters, models, and evaluation methods. The input parameters consist of a cluster structure (capturing the underlying network topology), covariate effects (how a cluster structure varies across subjects as a function of their covariates), random effects (introducing dependencies within-blocks and within-subjects), a network size (i.e., the total number of nodes n), a proportion design (i.e., the individual cluster sizes), a total number of visits (T), and a total number of subjects (K)

Lambiotte, Fornito, Ersche, & Bullmore, 2009) across many different species, such as the *C. elegans*, macaque and human brain networks (Achard et al., 2006; Bassett & Bullmore, 2009; Bullmore & Sporns, 2009; Hilgetag et al., 2000; Pan et al., 2010; Rubinov & Sporns, 2010; Sporns et al., 2004). However, there is also strong evidence that the cluster structure of a brain consists of a core of rich club nodes. For example, our previous work on the stochastic blockmodel analysis of the *C. elegans* brain network (Pavlovic et al., 2014) revealed a hybrid structure or “core-on-modules” structure, which deviated from a purely modular organization by the inclusion of densely inter- and intraconnected core blocks, which governed the worm's forward and backward motion. This kind of structure is hypothetically represented in the simulation by the core-modular topology. Finally, the weak cluster structure is an example of a topology where, on average, there is weak evidence for a three cluster solution as the connection profiles of Blocks 1 and 2 are identical. Such pathological cases may occur in noisy data sets. In *Simulation II*, however, we consider only the nonambiguous cluster structures with strong clustering evidence like the Het-Modular and core-modular structures so that there could be a fair comparison in the hypothesis testing procedures between Het-SBM and Het-Mixed-SBM.

TABLE 1 Block sizes for the three block size designs

Network sizes	Proportion design	Block sizes		
		n_1	n_2	n_3
$n = 50$	Balanced	17	17	16
	M. Unbalanced	30	15	5
	Unbalanced	35	10	5
$n = 100$	Balanced	34	34	32
	M. Unbalanced	60	30	10
	Unbalanced	70	20	10

Note: In the balanced design, all blocks have the same size while, in the mildly unbalanced (M. Unbalanced) and unbalanced designs, there is a systematic decrease in their number of nodes such that the last blocks are the smallest. In the M. Unbalanced design, the block sizes are mildly varying while, in the unbalanced case, the block sizes are changing more rapidly.

The rest of the simulation input parameters are self-explanatory in Figure 3. For example, in both simulations, the effect of the age covariate is set to zero, and the actual values for the covariates are randomly sampled. The random effect variances vary according to three values ($\sigma^2 \in \{0.5, 1, 2\}$), which are assumed to be constant across the entire cluster structure. The proportion designs specify the cluster sizes, and, among these, we consider three cases: balanced (clusters with similar sizes), mildly unbalanced (M. Unbalanced; clusters with a mild decrease in sizes), and unbalanced (clusters with a strong decrease in sizes). The exact number of nodes in each cluster for each proportion design is given in Table 1. The network sizes are $n \in \{50, 100\}$, the total number of subjects is $K \in \{20, 40, 80\}$, and the total number of visits is $T \in \{1, 3\}$ (with the exception of *Simulation II* where $T \in \{1\}$). For each such simulation scenario, we generate 100 Monte Carlo samples.

3.1.3 | Models.

For both simulations, we fit Het-SBM and Het-Mixed-SBM. In *Simulation I*, the models are fitted on the range of different cluster values $\hat{Q} \in \{2, 3, 4\}$ using the same initializations, while, in *Simulation II*, only the fitted parameters are used to benchmark the accuracy of the tests. For the data sets with multiple visits, we submit multiple visits to Het-SBM as independent subjects.

3.1.4 | Evaluation strategies.

The evaluation of the results in *Simulation I* is based on the accurate estimation of the optimal number of clusters and cluster labels. The comparison between the estimated clustering and the actual partition is carried out using the Adjusted Rand Index (ARI; Hubert & Arabie, 1985), which indicates the overall agreement between the selected best fits and the real clustering (with 1 denoting a perfect agreement). In *Simulation II*, we evaluate how well the parametric (Wald) and nonparametric (permutation) inference procedures control the false

positive rate (FPR) at a level of significance of 5%. The p -values of the Monte Carlo permutation test statistic are obtained by computing 1,000 permutations of the age covariate for each of the realizations.

3.2 | Simulation I: Results

For the data sets with the Het-Modular and core-modular cluster structures, Het-SBM and Het-Mixed-SBM both prove to be very accurate in the cluster label estimation and in selecting the optimal number of clusters ($ARI = 1$). This trend is noted for all the network sizes, subject totals, and random effect sizes (see Supplementary Information Figures S1–S8). This seems to indicate that, even under the presence of random effects, when the cluster profiles are well delineated, the choice between the fixed-effect or mixed-effect versions of the model is inconsequential in terms of clustering. However, as noted in Figure 4, this is not the case in the weak cluster structure data sets where the Het-SBM fits tend to be less accurate especially in the scenarios with small random effect variances (i.e., $\sigma^2 = 0.5$). For such values, Het-SBM struggles to pick on the between-subject variations in the connectivity profiles of Blocks 1 and 2, and it tends to underestimate the optimal number of clusters by one. It is interesting to point out that such effects are less pronounced when the network size is doubled, that is, when the individual clusters contain more nodes (see Figure 5). It is also worth pointing out that, in the cases of larger random effect variances, Het-SBM seems to be able to pick more accurately on the between-subject variations and

Weak cluster structure with 50 nodes and 1 visit

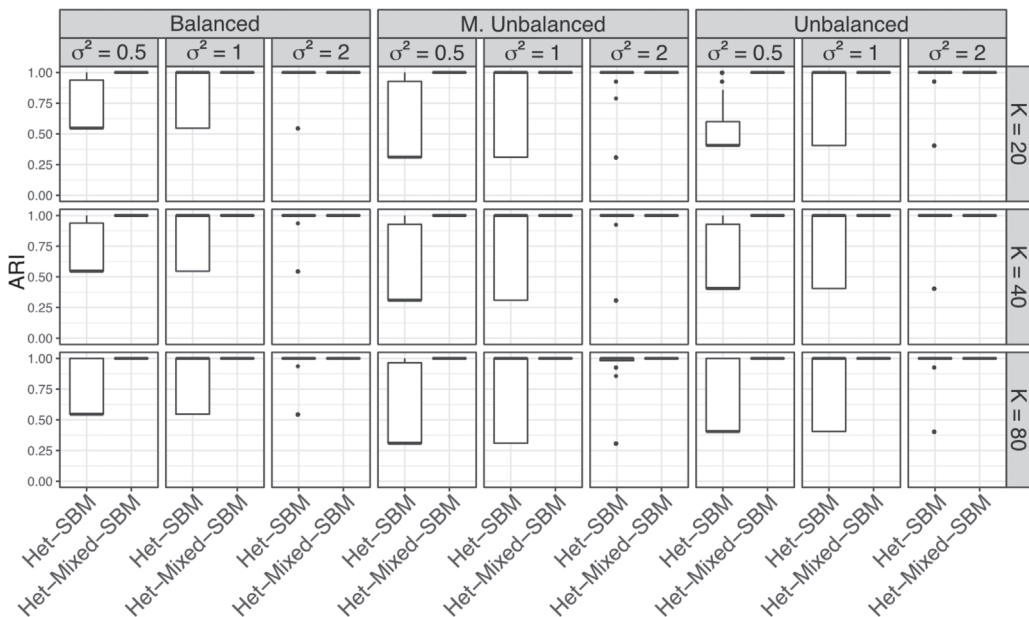


FIGURE 4 Box plots of ARI scores between the estimated and true cluster labels for the weak cluster structure, 50 nodes, and one visit. The x-axis shows two fitted models, Het-SBM and Het-Mixed-SBM, while the y-axis shows the distribution of their ARI scores. The rows show the subject totals, while the columns indicate the proportion designs and random effect variances

Weak cluster structure with 100 nodes and 1 visit

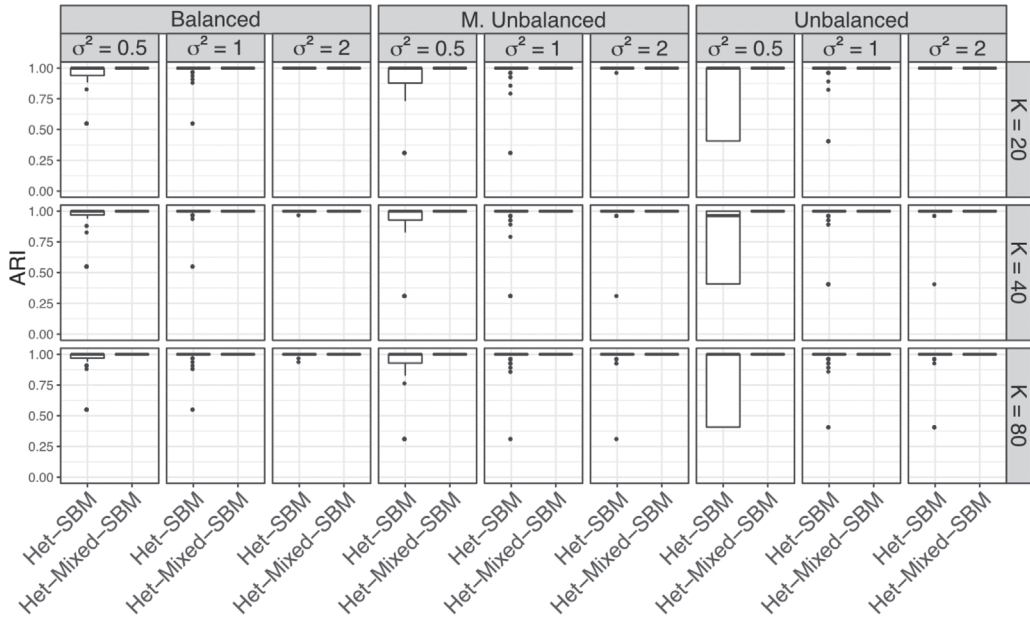


FIGURE 5 Box plots of ARI scores between the estimated and true cluster labels for the weak cluster structure, 100 nodes, and one visit. The x-axis shows two fitted models, Het-SBM and Het-Mixed-SBM, while the y-axis shows the distribution of their ARI scores. The rows show the subject totals, while the columns indicate the proportion designs and random effect variances

to retrieve the correct clustering. By contrast, Het-Mixed-SBM is accurate in all situations. This is because the model is able to use the block-level variance of random effect to guide the clustering.

3.3 | Simulation II: Results

The purpose of simulating the age effect under the null setting (i.e., $H_0: \beta_{ql}^{AGE} = 0$) is to evaluate the effectiveness of the Wald test and the permutation test based on the Wald statistics to control the Type I error rate at 5% significance. For convenience, we will hereafter refer to these procedures as the Wald test and permutation test, and we will base our discussion only on the simulation settings with the random-effect variances $\sigma^2 = 1$ as the trends observed are consistent with those observed in the other settings (whose results are given in Supplementary Information Appendix B).

Figures 6 and 7 show the observed false positive rates (FPRs) at a level of significance of 5% obtained from the tests on each block-specific age coefficient and how these results contrast across Het-SBM and Het-Mixed-SBM. Both figures display the results in the settings with 50 nodes, one visit and the variance of random effects $\sigma^2 = 1$ but with different cluster structures, namely, the Het-Modular cluster structure in Figure 6 and the core-modular cluster structure in Figure 7. In the case of the Het-Modular structure, we may expect a moderate bias on the age coefficients as some cluster-connection probabilities may be very close to zero or one. The Firth-type correction used in Het-SBM accounts for such a type of bias, but, for Het-Mixed-SBM, as this kind of bias prevention correction was not utilized, we may expect some degree of liberal behavior for the parametric test. As noted in Figure 6, the Wald test for Het-Mixed-SBM in Block (1, 2) displays

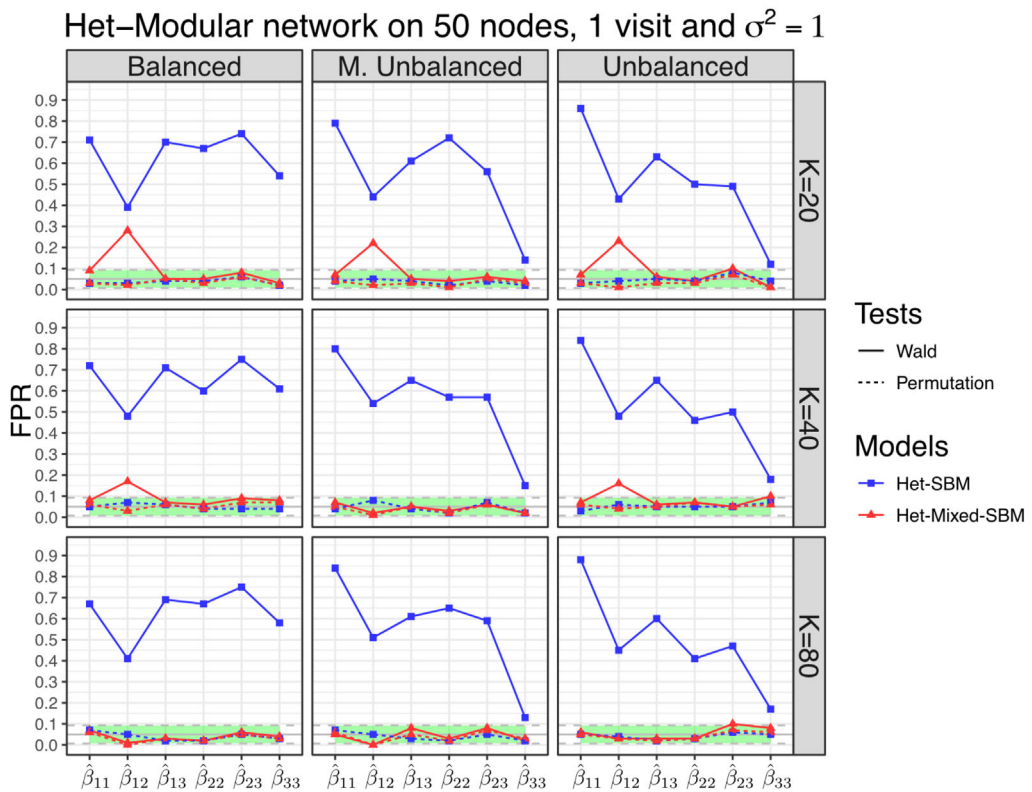


FIGURE 6 Observed false positive rates (FPRs) at a level of significance of 5% in the Het-Modular network with 50 nodes, one visit, and random effect variance $\sigma^2 = 1$. The columns of the plot show three proportion designs, while its rows show three different subject totals ($K \in \{20, 40, 80\}$). The x-axis shows the fitted cluster-level regression coefficients submitted to hypothesis testing for the effect of age (i.e., $H_0: \beta_{ql}^{AGE} = 0$) and the y-axis shows their corresponding observed FPR. The observed FPR are computed for the Wald test (solid line) and the permutation test based on the Wald statistics (dashed line). The results are shown for Het-SBM (blue line with square symbols) and Het-Mixed-SBM (red line with triangle symbols). The shaded green band represents the 95% confidence interval based on a normal approximation

exactly that behavior, which is not surprising as the input connection probability for that block was 0.01. Coupling such a small probability with a moderate number of subjects ($K = 20$) and a moderate number of edges introduces samples dominated by an excessive amount of zero-edges. With larger pulls of data (more edges and subjects), such effects seem to be ameliorated, resulting in better control of the FPR. For the core-modular structure in Figure 7, such effects seem to be absent and the Wald test appears to be valid in all scenarios. Similar trends are noted for other simulation settings (see Supplementary Information Figures S11–S14). In stark contrast to the results of Het-Mixed-SBM, the observed FPRs for the Wald test of Het-SBM are extremely inflated, suggesting complete unreliability of the test in the presence of random effects. One potential cause for this behavior lies in the assumption of edge independence within a block, which is no longer valid in the presence of random effects. In this particular setting, the fixed effects fail to explain the between-subject variation in the cluster-connectivity rates fully, and miss out on the dependence within the individual elements of a block structure introduced by the random effects. Notably, this trend is prevalent in all simulation settings and seems to moderately decrease only with a

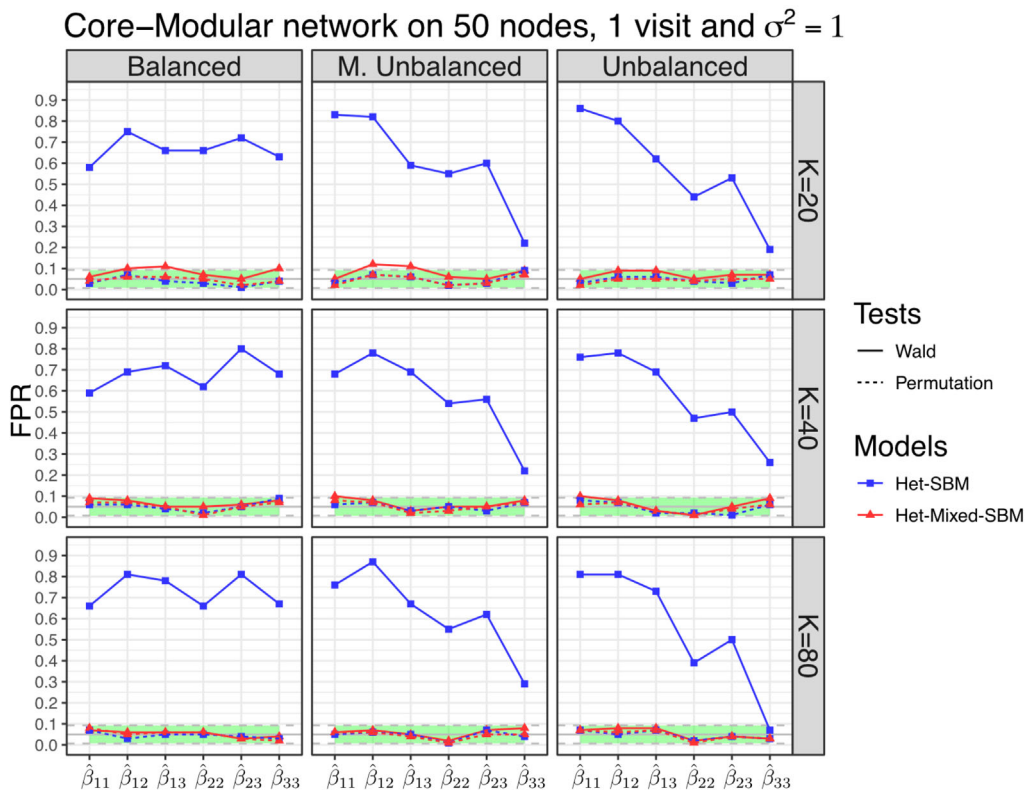


FIGURE 7 Observed false positive rates (FPRs) at a level of significance of 5% in the core-modular network with 50 nodes, one visit, and random effect variance $\sigma^2 = 1$. The columns of the plot show three proportion designs, while its rows show three different subject totals ($K \in \{20, 40, 80\}$). The x-axis shows the fitted cluster-level regression coefficients submitted to hypothesis testing for the effect of age (i.e., $H_0 : \beta_{qt}^{AGE} = 0$) and the y-axis shows their corresponding observed FPR. The observed FPRs are computed for the Wald test (solid line) and the permutation test based on the Wald statistics (dashed line). The results are shown for Het-SBM (blue line with square symbols) and Het-Mixed-SBM (red line with triangle symbols). The shaded green band represents the 95% confidence interval based on a normal approximation

smaller variance of the random effects (see Supplementary Information Figures S11 and S13 for $\sigma^2 = 0.5$). Finally, regarding the permutation test, it seems accurate in all simulation settings and for both models.

4 | HET-MIXED-SBM FIT TO MULTI-SUBJECT FUNCTIONAL CONNECTIVITY FMRI DATA

4.1 | Data

In this work, we consider a resting-state functional magnetic resonance imaging (fMRI) data set obtained from the Human Connectome Project (HCP), containing 100 unrelated and healthy subjects. Each subject fMRI scan has been fully preprocessed to account for potential confounding factors such as head motion and slice timing errors (Smith et al., 2013). Due to

excessive head motion and loss of signal in the frontal lobe, one subject has been removed from the study using the exclusion criterion recommended by Afyouni and Nichols (2018), thus yielding 99 subjects ($K = 99$). Next, each subject image has been further parcellated into 114 nonoverlapping regions of interests (i.e., $n = 114$) utilizing the methodological framework outlined in Thomas Yeo et al. (2011). Given that each region of interest contains several voxel-level time series, we use their mean time series for summarizing the functional activity in each node. However, even at this stage, the signal can still be contaminated by respiratory and cardiac confounds, which are also typically contained in the grand average time series (i.e., the global brain signal). Using the recommendations of Murphy and Fox (2017) and Li et al. (2019), the global mean signal is regressed out from each of the nodal time series. As noted in Figure 1, a functional connectivity matrix is obtained by correlating the nodal data in a pairwise fashion, and a binary network is derived from it by testing if each sample correlation is different from zero (i.e., $H_0 : r_{ijkt} = 0$ vs. $H_1 : r_{ijkt} > 0$). The negative correlation values bear very little neuroscientific information and, as such, are not considered in the analysis (Fornito, Zalesky, & Bullmore, 2016). The weakness of this approach, however, is that the fMRI time series are strongly autocorrelated (Woolrich, Ripley, Brady, & Smith, 2001), which produces a very poor estimate of the variance for Fisher Z scores used for the hypothesis tests, yielding strongly inflated false positives. In order to account for autocorrelation and cross-correlation across time series pairs, the approach termed “xDF” has been utilized (Afyouni et al., 2019). As shown in Afyouni et al. (2019), the xDF method produced much more accurate estimates than a naive approach. Using the xDF method, adjusted Z scores have been obtained, and the resulting p -values have been adjusted for multiple comparison error using a false discovery rate (Benjamini & Yekutieli, 2001). This part of the analysis has been completed using the Matlab xDF Toolbox (<https://github.com/asoroosh/xDF>, accessed on September 15, 2019; Afyouni et al., 2019). Finally, to obtain the binary adjacency matrix, any edge with an adjusted p -value lower than the nominal level (5%) has been set to one and otherwise to zero. Finally, for this data set, we only had one visit per subject, so $T = 1$ and all the subject covariates were not temporal.

In this data set, each subject is linked to the following covariates: age (29.20 ± 3.62), gender (+1 for a female with 54 subjects and -1 for a male with 45 subjects), and fluid intelligence score (IQ) (16.21 ± 4.62) given in terms of their means and standard deviations, respectively. Fluid intelligence has been evaluated using the Penn Progressive Matrices (PMAT) test in which the subjects have been tasked with predicting a correct pattern given a prior set of sequential patterns. The PMAT test consists of 24 items and the problems are arranged in order of increasing difficulty. The design matrix has been set up to contain a global intercept and the three covariates mentioned above (with age and IQ standardized). The three covariates have been chosen based on the fact that primary subject covariates like age and gender are known to have high explanatory power, especially for capturing differences between subjects (Dosenbach et al., 2010). Similar evidence has also been given in support of fluid intelligence (Dubois, Galdi, Paul, & Adolphs, 2018). It is worth mentioning, however, that even though the HCP data contain a wealth of behavioral scores, we have decided to select only the most standard covariates since there is very little empirical evidence that the available behavioral measures are informative of specific functional processes in the brain. Even less evidence exists about how informative these are to clustering. While we do not discount the potential research value of linking behavioral data to the functional connectivity data, we view this question as secondary to this work and beyond the scope of this article. Another possibility would be to use task-based fMRI data as a subject-specific covariate and to see

how these relate to the clustering. This would allow to study more directly the link between specific tasks and resting-state data. However, given that this model is one of the first steps toward linking covariates to clustering, more sophisticated analyses seem beyond the scope of this work.

4.2 | Clustering results

To this data, we fit Het-Mixed-SBM and Het-SBM using initializations based on the k-means clustering from the R package *amap* (Lucas, 2014) and the hierarchical clustering from the R package *stats* (R Core Team, 2017). For the k-means clustering, we perform 1,000 initializations for each subject and the grand average matrix while, for the hierarchical clustering, we perform only one initialization per each subject and the grand average since this method is deterministic. For demonstration purposes, the analysis only considers 10 clusters, and the ICL criterion in both models is used to select the best fits across initializations. The best fits for both models converge to an identical clustering solution, suggesting that the data do not contain noise levels, which would iron out connectivity profiles in some of the clusters as we have observed in our simulations. Therefore, the choice between the fixed-effect and mixed-effect modeling does not seem to impact the selected clustering fit. Although we do not observe differences in the clustering between Het-Mixed-SBM and Het-SBM in our real-data analysis, it is worth mentioning that this does not mean that this situation generalizes to all data sets. For example, it is possible to have similar mean connectivity profiles in some clusters but very different random effects and this would obscure the clustering with fixed effects (Het-SBM).

To understand the fit, in Figure 8, we list the individual nodes in each of the ten fitted blocks. The label of each node is sourced from the 17-network parcellation found in Thomas Yeo et al. (2011), which depicts typical neural activations while a subject is at rest. These labels are the central visual, control (A, B, and C), default (A, B, and C), dorsal attention (A and B), limbic (A and B), peripheral visual, salience/ventral (A and B), somatomotor (A and B), and temporal-parietal networks. *The central visual network* is engaged in conscious processing of visual stimuli, and it supports a depth of visual perception and conscious awareness of visual stimuli. The network comprises nodes from the striate and extrastriate cortex. *The control network* is supporting cognitive processes related to working memory, attention and decision making. In the parcellation, this network is split along three components labeled A, B, and C; the control A component consists of nodes in the temporal cortex, intraparietal sulcus, dorsal, lateral and lateral-ventral prefrontal cortex, and midcingulate cortex; the control B component consists of nodes in the temporal cortex, inferior parietal lobule, dorsal, lateral, lateral-ventral, medial-posterior, lateral-dorsal and lateral-ventral prefrontal cortex, and inferior parietal lobule; the control C component consists of nodes in the precuneus and posterior cingulate cortex. *The default mode network* is active during tasks such as daydreaming, future planning, accessing memories, and thinking about others or themselves. The network is split into three components: A, B, and C; the default A component consists of nodes in the inferior parietal lobule, dorsal prefrontal cortex, precuneus posterior cingulate cortex, medial prefrontal cortex, and temporal cortex; the default B component consists of nodes in the temporal cortex, anterior temporal lobe, inferior parietal lobule, dorsal, lateral, and ventral prefrontal cortex; the default C component consists of nodes in the precuneus, inferior parietal lobule, retrosplenial, and parahippocampal cortex. *The dorsal attention network* is linked to voluntary orienting of attention to external tasks. It is divided into two components: A and B; component A consists of nodes in the temporal and parietal occipital cortex and superior parietal lobule; component B consists of nodes in the temporal occipital cortex, postcentral cortex, frontal

Block 1 central visual somatomotor B salience/ventral attention B salience/ventral attention B control A control C central visual somatomotor B salience/ventral attention B control A control C	Block 2 central visual peripheral visual peripheral visual peripheral visual central visual peripheral visual peripheral visual peripheral visual	Block 3 somatomotor A somatomotor B somatomotor B somatomotor B salience/ventral attention A salience/ventral attention A salience/ventral attention A salience/ventral attention A salience/ventral attention A somatomotor A somatomotor B somatomotor B somatomotor B salience/ventral attention A salience/ventral attention A salience/ventral attention A salience/ventral attention A salience/ventral attention A salience/ventral attention A	Block 4 dorsal attention A dorsal attention A dorsal attention A dorsal attention B dorsal attention B dorsal attention B dorsal attention B dorsal attention A dorsal attention A dorsal attention A dorsal attention B dorsal attention B dorsal attention B dorsal attention B	Block 5 salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B salience/ventral attention B
Block 6 limbic A temporal parietal limbic A limbic B temporal parietal	Block 7 limbic B default A default A default A default A default B default B default B default B default B default A default A default A default A default B default B default B default B	Block 8 control A control A control A control A control A control A salience/ventral attention B control A control A control A control A	Block 9 control B control B control B control B control B control B control B control B control B control B control B	Block 10 control C default C default C default C control C default C default C default C

FIGURE 8 Network association of the Het-Mixed-SBM fit. Each block is given in terms of the 17-network parcellation from Thomas Yeo et al. (2011). The 17 networks are the central visual, control (A, B, and C), default (A, B, and C), dorsal attention (A and B), limbic (A and B), peripheral visual, salience/ventral (A and B), somatomotor (A and B), and temporal-parietal networks

eye fields, and precentral ventral cortex. *The limbic network* is linked to a range of emotional and behavioral responses. It is divided into components A and B; component A consists of nodes in the temporal pole; component B consists of nodes in the orbital frontal cortex. *The peripheral visual network* is supporting peripheral vision. It consists of nodes in the striate cortex, extrastriate inferior and extrastriate superior cortex. *The salience/ventral network* is linked to attention such as redirecting our attention to sudden stimuli as well as for deciding which such stimuli are worthy of attention. It is divided into components A and B; component A consists of nodes in the parietal operculum, frontal operculum, insula, parietal medial cortex, frontal medial cortex, and precentral cortex; component B consists of nodes in the cingulate anterior cortex, inferior parietal lobule, lateral prefrontal cortex, medial posterior prefrontal cortex, and dorsal prefrontal cortex. *The somatomotor network* is linked to the coordination of motor tasks. It consists of components A and B; the component A consists of nodes in the primary somatosensory cortex; the component B consists of nodes in the insula, central cortex, secondary somatosensory cortex, and auditory cortex. *The temporal-parietal network* contains regions in the temporal parietal cortex.

Next, we summarize each block according to its functional specification (see Figure 9). Block 1 contains regions from five networks: the central visual, somatomotor B, salience/ventral attention B, control A and C networks. Given the diversity of its network associations, it is not easy to attach a unique functional label. Nevertheless, each region in this block seems to be very weakly connected with the other regions in this block and the regions in the other blocks, which can

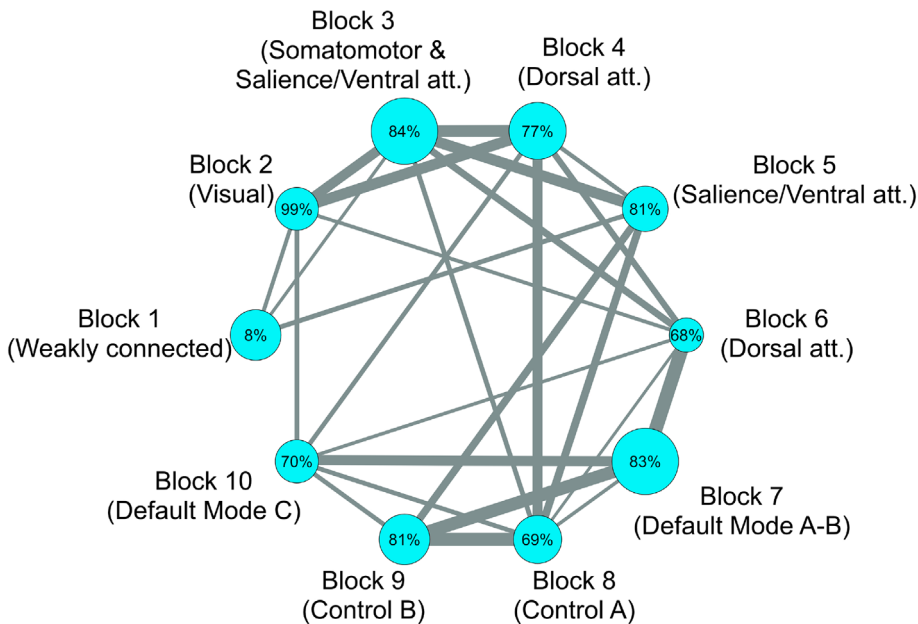


FIGURE 9 Average connectivity matrix (for mean age, gender, and IQ) thresholded at 10% connectivity rate. The clusters are plotted along a circle, and their relative sizes correspond to their total number of nodes. Within-cluster connections are given in percent, and between-cluster links are provided in terms of their edges whose thicknesses indicate their relative strength of association. Each block is also awarded an approximate functional label

explain why they have been clustered together. For this reason, we refer to this block as “weakly connected.” By contrast, the remaining blocks are more straightforward to classify functionally. Indeed, Block 2 can be categorized as “visual” as it contains elements from the central and peripheral visual networks, Block 3 as “somatomotor and salience/ventral attention,” Block 4 as “dorsal attention,” Block 5 as “salience/ventral attention B,” Block 6 as “limbic and temporal-parietal,” Block 7 as “default mode A-B,” Block 8 as “control A,” Block 9 as “control B,” and Block 10 as “default mode C.”

Figure 10 shows the average connectivity matrix for mean age, gender, and IQ. The within-cluster connectivity of Block 1 is fragile and weaker than several of its connections with other blocks (e.g., the connectivity between Block 1 and Block 5), suggesting a so-called “disassortative mixing” pattern. The nodes in this block appear to be classified together based on their substantial dissimilarity in connectivity profiles to the nodes in the rest of the networks. This departure from the main clustering patterns may potentially explain its functional heterogeneity, which was identified earlier. It is also interesting to point out that Blocks 2 (visual), 3 (somatomotor and salience/ventral attention), and 4 (dorsal attention) exhibit relatively high within-cluster connectivities but also relatively strong connections between one another. The high within-cluster connectivity is generally expected since regions in close anatomical proximity are involved in local processes. The relatively high connectivity between these clusters suggests an interplay between visual, somatomotor, and attention tasks. Although the components of salience/ventral attention network are split between Blocks 3 and 5 due to their differing connectivity profiles (e.g., Block (3, 9) and Block (5, 9)), these blocks maintain relatively high mutual connections, which are expected within the same functional network. Blocks 6

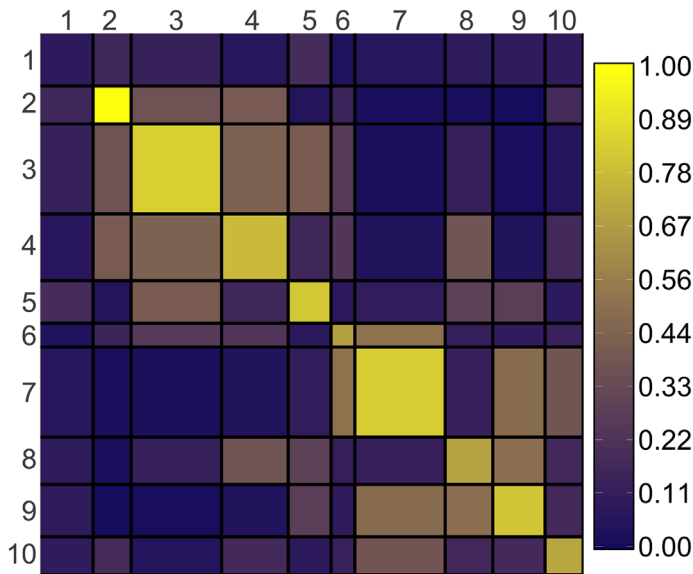


FIGURE 10 Average connectivity matrix for mean age, gender, and IQ. The emerging cluster structure seems to diverge from a classical modular organization and shows evidence of high connectivity patterns between some blocks (e.g., between Blocks 2, 3, and 4, between Blocks 6 and 7, and between Blocks 8 and 10). The blocks are defined according to their connectivity profile and even though Blocks 3 and 4 appear to have similar within- and between-cluster probabilities, they are split into separate blocks on the grounds of their different connections to Blocks 5 and 8. An exception from the overall structural pattern in the cluster structure is Block 1, which has more connections to the other blocks in the network than within its cluster. This pattern is generally known as a disassortative mixing

(limbic and temporal-parietal) and 7 (default mode A-B) also exhibit strong mutual connections, while the component C of the default mode network in Block 10 is not as connected to Block 6 as Block 7. Finally, the components A and B of the control network are split into Blocks 8 and 9, respectively.

4.3 | Inference results

In Figure 11, we show the uncorrected p -values (in $-\log_{10}$ scale) thresholded at 5% significance for the null hypothesis of no effect for each covariate (age, gender, and IQ). The results are shown for four testing procedures based on the combination of the two models (Het-SBM and Het-Mixed-SBM) and two tests (Wald and permutation tests). The Wald test for Het-SBM exhibits a much higher number of significant blocks than the three other testing procedures. This behavior seems to be consistent with the results shown in *Simulation II* (see Section 3.3), where we have observed a very strong liberal control of the FPR for the Wald test of Het-SBM. In general, the results of the three other testing procedures seem relatively similar.

In Figure 12, we show the familywise error rate corrected p -values (in $-\log_{10}$ scale) thresholded at 5% significance for the null hypothesis of no effect for each covariate (age, gender, and IQ) and each testing procedure. As with the uncorrected p -values, the Wald test for Het-SBM

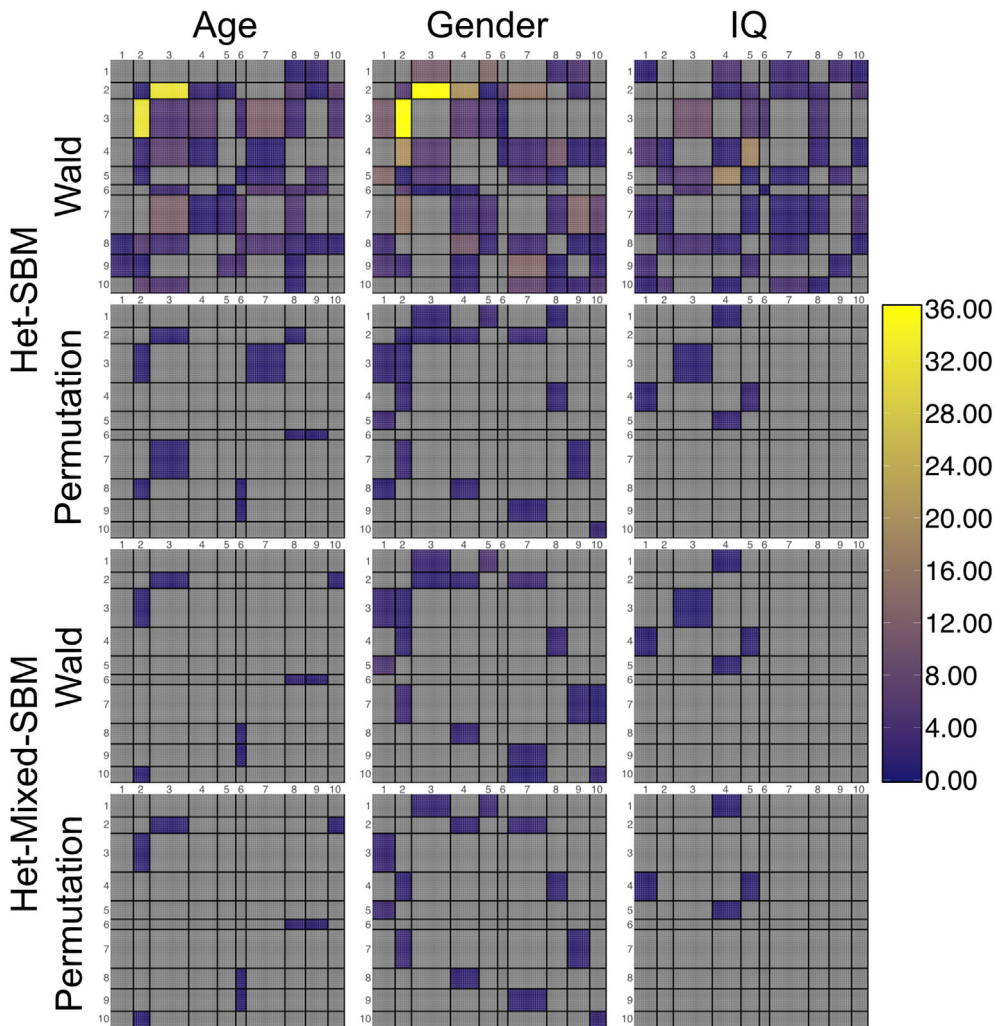


FIGURE 11 Uncorrected significant p -values for the null hypothesis of no effect for each covariate (age, gender, and IQ). The results are given for Het-SBM and Het-Mixed-SBM and in terms of the Wald and the permutation tests. The p -values are stated in terms of $-\log_{10}(p\text{-values})$ (e.g., the legend score of 36 corresponds to a p -value of 10^{-36}) and the nonsignificant p -values at 5% significance are given in gray

exhibits a much higher number of significant blocks than the three other testing procedures. The permutation test for Het-SBM declared two significant results that occur for age and gender in Block (2,3) (between the visual and somatomotor and salience/ventral attention blocks). In contrast to this, the Wald and permutation tests for Het-Mixed-SBM both consistently declared a single significant result that occurs for gender in Block (1,5) (between the weakly connected and salience/ventral attention blocks). As shown in Nichols and Hayasaka (2003), one potential explanation for this discrepancy between the two models is that the permutation test based on the maximum statistic is sensitive to situations where the statistics are not pivotal. As the statistics in Het-SBM do not account for random effects, their null distributions are likely to vary across blocks, especially if the variance of the random effects is changing across blocks. In such cases,

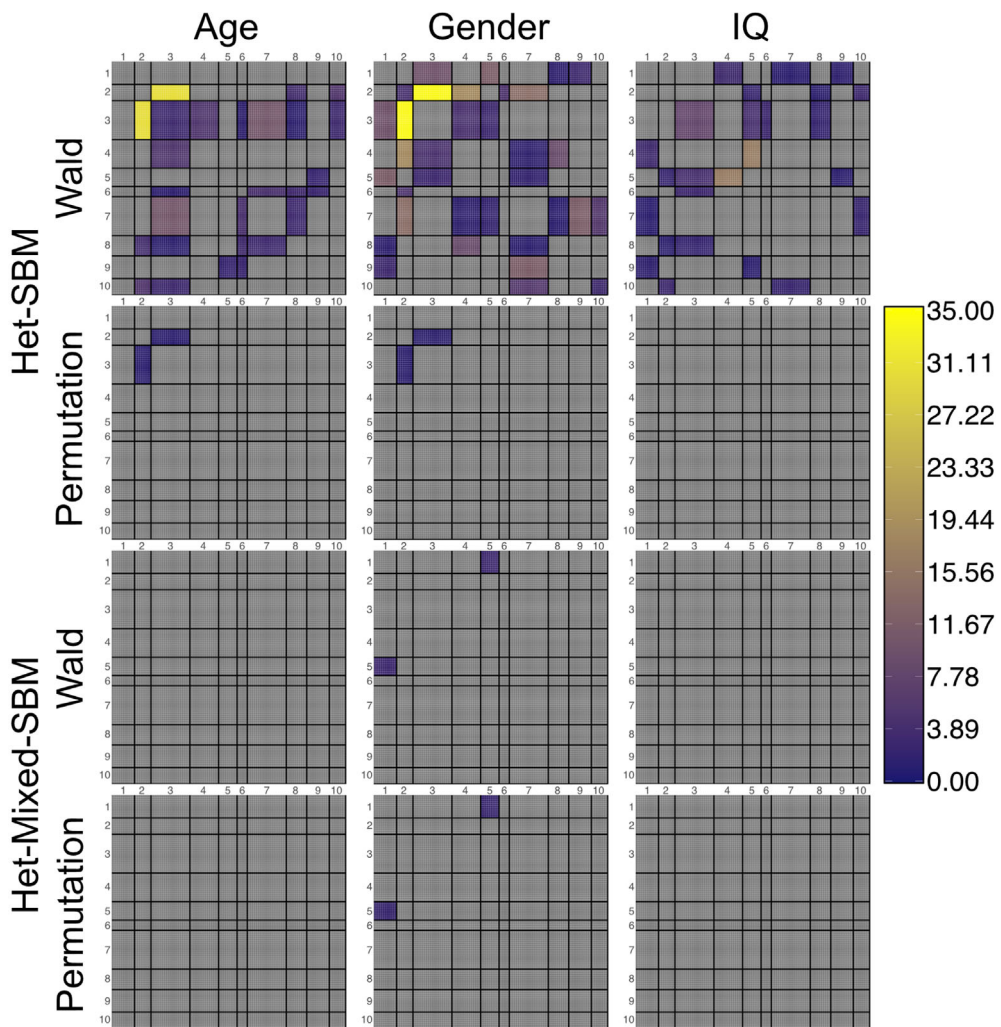


FIGURE 12 Familywise error rate corrected significant p -values for the null hypothesis of no effect for each covariate (age, gender, and IQ). The results are given for Het-SBM and Het-Mixed-SBM and in terms of the Wald and the permutation tests. The p -values are stated in terms of $-\log_{10}(p\text{-values})$ (e.g., the legend score of 35 corresponds to a p -value of 10^{-35}) and the nonsignificant p -values at 5% significance are given in gray. For the Wald test, the familywise error rate corrected p -values were obtained using a Bonferroni correction. For the permutation test, the familywise error rate corrected p -values were obtained based on the null distribution of the maximum statistics

the null distribution of the maximum statistic is likely to be driven by the blocks that bear the most familywise error risks. As a consequence of that, some blocks will be more penalized than others. On the contrary, as Het-Mixed-SBM accounts for some random effects, we may expect the statistics to be more pivotal. This means that the contribution of each block to the null distribution of the maximum statistic is likely to be more homogeneous, penalizing blocks more equally. It is worth mentioning that, based on this explanation, both results are valid even if they do not exhibit the same significant blocks because they are both controlling the familywise error rate.

5 | DISCUSSION

In this article, we proposed a novel multi-subject model, termed Het-Mixed-SBM, which can be used for the analysis of multi-subject binary network data, allowing for random effects per subjects and blocks.

In *Simulation I*, we validated the clustering ability of the model on a range of different cluster structures, and we have compared its results with the Het-SBM fits. In the instances of clearly delineated cluster structures in which the blocks show starkly segregated connectivity profiles (i.e., the Het-Modular and Core-Modular structures), both models converged to the correct clustering solutions. Nevertheless, in the cluster structure with merged connectivity profiles, such as Blocks 1 and 2 in the weak topology, the variance of the random effects can be used by Het-Mixed-SBM to correctly separate Blocks 1 and 2. Interestingly, for small variations of the random effects, it seemed hard for Het-SBM to split Blocks 1 and 2 across the subjects, and the model tended to underestimate the optimal number of clusters. However, when the random effect variances were larger, Het-SBM yielded an accurate clustering. The overall conclusion of *Simulation I* is that, for cluster structures with strong clustering evidence, we can expect a similar performance between the two models. Still, for cluster structures with weaker evidence, it might be possible to get erroneous solutions with the fixed-effects model (Het-SBM). In combination with the conclusions of Pavlovic et al. (2020) where it was shown that the fixed-effect covariates might be informative of the clustering, it seems that the random effects can also be useful in guiding the model toward identifying the correct clustering solution in multi-subject networks. As such, it may not be appropriate to do the clustering separately from the logistic regression or logistic-mixed regression.

Having established in *Simulation I* the influence of random effects on the clustering, in *Simulation II*, we focused on validating the inference procedures based on the parametric (Wald) and nonparametric (permutation) tests in Het-SBM and Het-Mixed-SBM. As previously anticipated, the Wald test in Het-SBM showed highly inflated FPRs in the presence of random effects in the data, and such results seem to improve with the permutation testing. By contrast, the Wald and permutation testing in Het-Mixed-SBM showed more consistent results. The only departure from this trend appears in the particular block of the Het-Modular structure whose low connectivity rate introduced samples with hugely disproportionate ratios of zeros to ones, which are known to yield highly biased estimates and inflated FPRs for the Wald test. To remove this type of bias, one solution would be to use Firth-type estimators like in Het-SBM. Nevertheless, this is not straightforward due to the presence of random effects in the model. From a purely practical viewpoint, in such circumstances, it may be advisable to resort to a more straightforward alternative like permutation testing. Nevertheless, for probability rates that are away from the zero/one boundaries, it may be simpler to use the Wald test.

In the analysis of the HCP data set, we obtained identical clustering fits for Het-SBM and Het-Mixed-SBM. The clustering fit highlighted functionally meaningful blocks. The inference results based on the Wald and permutation testing in Het-SBM showed widely different results. These results echo the conclusions of *Simulation II*, which suggested a liberal behavior of the Wald test with Het-SBM. By contrast, there seems to be a high degree of agreement between the Wald and the permutation testing in Het-Mixed-SBM. These results seem to reinforce the notion that it is important to account for random effects to better model the data. A shortcoming of our real-data analysis is that we fixed the total number of clusters to 10 to ensure that the blocks are reasonably sized for realistic inferences and to avoid potential small sample issues. Further

analyses using higher numbers of clusters would be useful to uncover finer-grained fits that might be more informative.

One shortcoming of Het-Mixed-SBM is that it only considers a random intercept per block as random effects. While this is an improvement upon the classic Het-SBM which does not model any form of within-subject dependencies, it assumes that, for a given block, the within-subject dependencies are the same between all the within-subject connections. This assumption may be too restrictive for repeated measures data as we may expect the dependencies to be more complicated. For example, in the case of longitudinal networks, we may expect the dependencies to decrease with time, which would not be modeled with our current version of Het-Mixed-SBM. One potential remedy in modeling more complex types of within-subject dependencies is to include additional random effects to the model. For example, in the case of longitudinal networks, adding a random effect of time would allow to model dependencies that also vary between time points. Similar considerations might be given to data that combined different imaging modalities (e.g., a combination of diffusion tensor imaging and resting-state fMRI).

Finally, as part of our future work, we will also consider a potentially useful simplification of our model that may speed up computation while accounting for subject-level randomness. Specifically, we intend to consider a model that combines “homogeneous random effects” (per-subject random intercepts common to the entire network) with “heterogeneous fixed effects” regression for each block element. We expect this model to have a similar clustering potential as the Het-Mixed-SBM but with a much reduced computational burden.

ORCID

Dragana M. Pavlović  <https://orcid.org/0000-0001-5083-8731>

REFERENCES

- Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, 26(1), 63–72.
- Afyouni, S., & Nichols, T. E. (2018). Insight and inference for DVARS. *Neuroimage*, 172, 291–312.
- Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, 199, 609–625.
- Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of machine learning research*, 9, 1981–2014.
- Ambrose, C., & Matias, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1), 3–35.
- Bassett, D. S., & Bullmore, E. T. (2009). Human brain networks in health and disease. *Current Opinion in Neurology*, 22(4), 340.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188.
- Bickel, P., Choi, D., Chang, X., & Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4), 1922–1943.
- Biernacki, C., Celeux, G., Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719–725.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.

- Choi, D. S., Wolfe, P. J., & Airoldi, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika*, 99(2), 273–284.
- Côme, E., & Latouche, P. (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling*, 15(6), 564–589.
- Daudin, J. J., Picard, F., & Robin, S. (2008). A mixture model for random graphs. *Statistics and computing*, 18(2), 173–183.
- Demidenko, E. (2004). *Mixed models: Theory and applications Wiley series in probability and statistics* (). Hoboken, NJ: Wiley.
- Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., ... Barnes, K. A. (2010). Prediction of individual brain maturity using fMRI. *Science*, 329(5997), 1358–1361.
- Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170284.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38.
- Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of brain network analysis*. Cambridge, MA: Academic Press.
- Gao, C., Lu, Y., & Zhou, H. H. (2015). Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6), 2624–2652.
- Good, P. (2000). *Permutation tests*. New York, NY: Springer.
- Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Scannell, J. W., & Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355(1393), 91–110.
- Hinne, M., Ekman, M., Janssen, R. J., Heskes, T., & van Gerven, M. A. (2015). Probabilistic clustering of the human connectome identifies communities and hubs. *PloS One*, 10(1), e0117179.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218.
- Latouche, P., Birmele, E., & Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1), 93–115.
- Latouche, P., Birmelé, E., & Ambroise, C. (2014). *Overlapping clustering methods for networks*. Boca Raton, FL: Chapman & Hall/CRC.
- Latouche, P., Birmelé, E., & Ambroise, C. (2011). Overlapping stochastic block models with application to the french political blogosphere. *The Annals of Applied Statistics*, 5(1), 309–336.
- Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3), 325–335.
- Li, J., Kong, R., Liegeois, R., Orban, C., Tan, Y., Sun, N., ... Yeo, B. T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage*, 196, 126–141.
- Liu, Q., & Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika*, 81(3), 624–629.
- Lucas A. 2014. AMAP: Another multidimensional analysis package. R package version 0.8-14. Retrieved from <https://CRAN.R-project.org/package=amap>
- Mariadassou, M., Robin, S., & Vacher, C. (2010). Uncovering latent structure in valued graphs: A variational approach. *The Annals of Applied Statistics*, 4(2), 715–742.
- Matias, C., & Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4), 1119–1141.
- Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., & Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3, 37. doi: 10.3389/neuro.11.037.2009.
- Moyer D, Gutman B, Prasad G, Faskowitz J, Ver Steeg G., & Thompson P. 2015. *Blockmodels for connectome analysis*. Paper presented at: Proceedings of the 11th International Symposium on Medical Information Processing and Analysis (vol. 9681, p. 96810A). International Society for Optics and Photonics.
- Murphy, K., & Fox, M. D. (2017). Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *Neuroimage*, 154, 169–173.
- Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, 12(5), 419–446.

- Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455), 1077–1087.
- Olhede, S. C., & Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41), 14722–14727.
- Pan, R. K., Chatterjee, N., & Sinha, S. (2010). Mesoscopic organization reveals the constraints governing *Caenorhabditis elegans* nervous system. *PloS One*, 5(2), e9240.
- Pavlovic DM. 2015. *Generalised stochastic blockmodels and their applications in the analysis of brain networks* (PhD thesis).
- Pavlovic, D. M., Guillaume, B. L., Towlson, E. K., Kuek, N. M., Afyouni, S., Vertes, P. E., Yeo, B.T.T., & Nichols, T. E. (2020). Multi-subject stochastic blockmodels for adaptive analysis of individual differences in human brain network cluster structure. *NeuroImage*, 116611.
- Pavlovic, D. M., Vértes, P. E., Bullmore, E. T., Schafer, W. R., & Nichols, T. E. (2014). Stochastic blockmodeling of the modules and core of the *Caenorhabditis elegans* connectome. *PLoS One*, 9(7), e97584.
- Picard, F., Miele, V., Daudin, J. J., Cottret, L., & Robin, S. (2009). *Deciphering the connectivity structure of biological networks using MixNet*. In *BMC bioinformatics* (Vol. 10(S6), p. S17).
- Piessens R, Doncker-Kapenga D, Überhuber C, & Kahaner D. 1983. QUADPACK: A subroutine package for automatic integration. status: published. 301p.
- Potter, D. M. (2005). A permutation test for inference in logistic regression with small-and moderate-sized data sets. *Statistics in Medicine*, 24(5), 693–708.
- R Core Team. (2015, Retrieved from). *R: A language and environment for statistical computing*. Vienna, Austria: Computing. R Foundation for Statistical Computing. <http://www.R-project.org/>
- R Core Team. (2017, Retrieved from). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robinson, L. F., Atlas, L. Y., & Wager, T. D. (2015). Dynamic functional connectivity using state-based dynamic community structure: Method and application to opioid analgesia. *NeuroImage*, 108, 274–291.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3), 1059–1069.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., ... Kelly, M. (2013). Resting-state fMRI in the human connectome project. *Neuroimage*, 80, 144–168.
- Snijders, T. A. B., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1), 75–100.
- Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9), 418–425.
- Thomas Yeo, B., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165.
- Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., ... Della-Penna, S. (2012). The human connectome project: A data acquisition perspective. *Neuroimage*, 62(4), 2222–2231.
- Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment* (Vol. 279). Hoboken, NJ: John Wiley & Sons.
- Wolfe PJ, Olhede SC. 2013. *Nonparametric graphon estimation*. arXiv preprint arXiv:13095936.
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage*, 14(6), 1370–1386.
- Zanghi, H., Ambroise, C., & Miele, V. (2008). Fast online graph clustering via Erdős-Rényi mixture. *Pattern Recognition*, 41(12), 3592–3599.
- Zanghi, H., Volant, S., & Ambroise, C. (2010). Clustering based on random graph model embedding vertex features. *Pattern Recognition Letters*, 31(9), 830–836.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Pavlović DM, Guillaume BR, Afyouni S, Nichols TE.

Multi-subject stochastic blockmodels with mixed effects for adaptive analysis of individual differences in human brain network cluster structure. *Statistica Neerlandica*.

2020;74:363–396. <https://doi.org/10.1111/stan.12219>

APPENDIX A. PARAMETER ESTIMATION IN HET-MIXED-SBM

For the clarity of the subsequent discussions and easy referencing, we provide a list of integrals $I_{qlk1} - I_{qlk6}$ that will be used in the estimating equations of β and σ^2 . Thus, we have

$$I_{qlk1} = \int_{-\infty}^{+\infty} e^{h_{qlk}} dr_{qlk}, \quad (A1)$$

$$I_{qlk2} = \frac{\partial I_{qlk1}}{\partial \sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 dr_{qlk}, \quad (A2)$$

$$I_{qlk3} = \frac{\partial I_{qlk1}}{\partial \beta_{ql}} = \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} dr_{qlk}, \quad (A3)$$

$$I_{qlk4} = \frac{\partial I_{qlk2}}{\partial \sigma_{ql}^2} = \frac{1}{4\sigma_{ql}^8} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^4 dr_{qlk} - \frac{1}{\sigma_{ql}^6} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 dr_{qlk}, \quad (A4)$$

$$\begin{aligned} I_{qlk5} = \frac{\partial \mathbf{I}_{qlk3}^\top}{\partial \beta_{ql}} &= \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right)^2 \sum_{t=1}^T \sum_{t'=1}^T \mathbf{d}_{kt} \mathbf{d}_{kt'}^\top \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} \pi_{qlkt'} dr_{qlk} \\ &+ \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \mathbf{d}_{kt}^\top \int_{-\infty}^{+\infty} e^{h_{qlk}} \pi_{qlkt} (1 - \pi_{qlkt}) dr_{qlk}, \end{aligned} \quad (A5)$$

$$I_{qlk6} = \frac{\partial I_{qlk3}}{\partial \sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4} \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk}} r_{qlk}^2 \pi_{qlkt} dr_{qlk}. \quad (A6)$$

Taking the partial derivatives of the variational bound $\mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$ (Equation (10)) with respect to β_{ql} and σ_{ql}^2 , for an individual block (q, l) , we can form a score vector $\mathbf{U}(\beta_{ql}, \sigma_{ql}^2)$ as a $(P+1) \times 1$ vector of first derivatives $\mathbf{U}(\beta_{ql}, \sigma_{ql}^2) = (\mathbf{U}(\beta_{ql})^\top, U(\sigma_{ql}^2)^\top)^\top$ whose equations are

given as

$$\mathbf{U}(\boldsymbol{\beta}_{ql}) = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt} \mathbf{d}_{kt} + \sum_{k=1}^K \frac{\mathbf{I}_{qlk3}}{I_{qlk1}}, \quad (\text{A7})$$

$$U(\sigma_{ql}^2) = -\frac{K}{2\sigma_{ql}^2} + \sum_{k=1}^K \frac{I_{qlk2}}{I_{qlk1}}. \quad (\text{A8})$$

Similarly, the negative second-order partial derivatives of the variational bound yield the observed Fisher Information matrix $\mathcal{F}(\boldsymbol{\beta}, \sigma^2)$, which appears in this model as a block diagonal matrix whose overall dimension is $(P+1)Q(Q+1)/2 \times (P+1)Q(Q+1)/2$. In particular, each diagonal block element is noted as $\mathcal{F}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2)$ and is a $(P+1) \times (P+1)$ matrix that can be represented as

$$\mathcal{F}(\boldsymbol{\beta}_{ql}, \sigma_{ql}^2) = \left[\begin{array}{c|c} \mathcal{F}_1(\boldsymbol{\beta}_{ql}) & \mathcal{F}_2(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) \\ \hline \mathcal{F}_2^T(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) & \mathcal{I}_3(\sigma_{ql}^2) \end{array} \right], \quad (\text{A9})$$

such that

$$\mathcal{F}_1(\boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}_{ql} \partial \boldsymbol{\beta}_{ql}^T} = -\sum_{k=1}^K \left[\frac{I_{qlk5}}{I_{qlk1}} - \frac{I_{qlk3} \mathbf{I}_{qlk3}^T}{I_{qlk1}^2} \right], \quad (\text{A10})$$

$$\mathcal{F}_2(\sigma_{ql}^2, \boldsymbol{\beta}_{ql}) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma_{ql}^2 \partial \boldsymbol{\beta}_{ql}} = -\sum_{k=1}^K \left[\frac{I_{qlk6}}{I_{qlk1}} - \frac{I_{qlk3} I_{qlk2}}{I_{qlk1}^2} \right], \quad (\text{A11})$$

$$\mathcal{I}_3(\sigma_{ql}^2) \equiv -\frac{\partial^2 \mathcal{J}(f^*(\mathbf{z}; \boldsymbol{\tau}); \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)}{\partial \sigma_{ql}^2 \partial \sigma_{ql}^2} = -\frac{K}{2\sigma_{ql}^4} - \sum_{k=1}^K \left[\frac{I_{qlk4}}{I_{qlk1}} - \frac{I_{qlk2}^2}{I_{qlk1}^2} \right]. \quad (\text{A12})$$

APPENDIX B. STARTING VALUES FOR THE VARIANCES OF THE RANDOM EFFECTS

In this section, we describe a procedure for obtaining the initial values of σ^2 . The goal is to first obtain an estimate of the random effects or the deviation of a subject specific intercept from the population intercept, then estimate their variances and finally construct the initial value estimates of σ^2 . For the k th subject and the (q, l) th block, the goal is to maximize h_{qlk} with respect to r_{qlk} . Dropping the terms which depend on σ_{ql}^2 , we have

$$h_{qlk}(r_{qlk}) = \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt} r_{qlk} - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \log(1 + e^{\mathbf{d}_{kt}^T \boldsymbol{\beta}_{ql} + r_{qlk}}). \quad (\text{B1})$$

Taking the derivatives with respect to r_{qlk} , gives $U(r_{qlk})$

$$U(r_{qlk}) = \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql}}{1 + e^{\mathbf{d}_{kt}^\top \beta_{ql} + r_{qlk}}} - \frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt}, \quad (\text{B2})$$

and setting $U(r_{qlk}) = 0$, we get

$$\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql}}{1 + e^{\mathbf{d}_{kt}^\top \beta_{ql} + r_{qlk}}} = \frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt}. \quad (\text{B3})$$

Using the substitution $\omega_{qlk} = e^{r_{qlk}}$, we obtain

$$\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql}}{1 + \omega_{qlk} e^{\mathbf{d}_{kt}^\top \beta_{ql}}} = \frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt}, \quad (\text{B4})$$

and taking the derivatives with respect to ω_{qlk} gives

$$-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql} e^{\mathbf{d}_{kt}^\top \beta_{ql}}}{(1 + \omega_{qlk} e^{\mathbf{d}_{kt}^\top \beta_{ql}})^2}. \quad (\text{B5})$$

An estimate of ω_{qlk} can then be found via the Newton–Raphson updating equations

$$\omega_{qlk}^{(m+1)} = \omega_{qlk}^{(m)} + \frac{\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt} - \frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql}}{1 + \omega_{qlk}^{(m)} e^{\mathbf{d}_{kt}^\top \beta_{ql}}}}{\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \frac{\gamma_{ijql} e^{\mathbf{d}_{kt}^\top \beta_{ql}}}{(1 + \omega_{qlk}^{(m)} e^{\mathbf{d}_{kt}^\top \beta_{ql}})^2}}}. \quad (\text{B6})$$

For the starting point ω_{qlk}^0 , we take advantage of the relationship between the arithmetic and harmonic mean so that

$$\sum_{t=1}^T \frac{1}{1 + e^{\mathbf{d}_{kt}^\top \beta_{ql}} \omega_{qlk}^0} \geq \frac{T^2}{T + \omega_{qlk}^0 \sum_{t=1}^T e^{\mathbf{d}_{kt}^\top \beta_{ql}}}. \quad (\text{B7})$$

Substituting (B7) into (B4), we can estimate the starting point ω_{qlk}^0 as

$$\omega_{qlk}^0 = \frac{\frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt}}{\sum_{t=1}^T e^{\mathbf{d}_{kt}^\top \beta_{ql}} \left(\frac{T}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt} \right)}. \quad (\text{B8})$$

For the solution of the Newton–Raphson step $\omega_{qlk}^{(*)}$, we take the transformation $r_{qlk} = \log(\omega_{qlk}^{(*)})$. After obtaining an estimate for every subject k , we get a sequence $\{r_{ql1}, \dots, r_{qlk}\}$. Then the goal is to use this sequence to estimate σ_{ql}^2 . For this, we need $\text{Var}(r_{qlk}) = v_{qlk}$, which is just

$$v_{qlk} = \frac{1}{-\frac{\partial}{\partial r_{qlk}} U(r_{qlk})} = \frac{1}{\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \frac{e^{\mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql}}}{1 + e^{\mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql}}}}, \quad (\text{B9})$$

and finally the estimate of a starting point is

$$\hat{\sigma}_{ql}^2 = \frac{1}{K} \sum_{k=1}^K (r_{qlk}^2 - v_{qlk}). \quad (\text{B10})$$

APPENDIX C. REPARAMETRIZATION OF THE INTEGRALS IN THE HET-MIXED-SBM

At each step of the Newton–Raphson algorithm, we require a numerical approximation of the six integrals (Equations (A1)–(A6)). In practice, such integrals can be reasonably well estimated with the adaptive Gauss–Hermite quadrature approximation (Lesaffre & Spiessens, 2001; Liu & Pierce, 1994) whose implementation in R is available via the function `integrate` (Piessens, Doncker-Kapenga, Überhuber, & Kahaner, 1983; R Core Team, 2015). Although this computational strategy offers relatively quick and accurate approximations, in a few examples, we encountered some instabilities and numerical issues. First, some integrals were evaluated as zeros, which caused the variational bound to diverge to $-\infty$. The main reason for this was that the maximal value attained by the function h_{qlk} was very small (e.g., $-1,000$), making the integrand numerically equal to zero over the whole domain of integration. To treat this point of numerical instability, we can add an offset value c_{qlk} to the exponent h_{qlk} . In particular, this offset value can be chosen to be the negative maximum of h_{qlk} , taken with respect to r_{qlk} , while iterative values of $\boldsymbol{\beta}_{ql}^{(m)}$ and $\sigma_{ql}^{2(m)}$ are assumed to be fixed. Thus, we have

$$c_{qlk} = - \max_{r_{qlk} \in [-1,000, +1,000]} \left(h_{qlk}(\boldsymbol{\beta}_{ql}^{(m)}, \sigma_{ql}^{2(m)}) \right). \quad (\text{C1})$$

Using this, we can now write the integral $\int_{-\infty}^{+\infty} e^{h_{qlk}} dr_{qlk} = e^{-c_{qlk}} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} dr_{qlk}$ and the edge based component of the variational bound is given as

$$\begin{aligned} \mathbb{E}(\log f(\mathbf{x}|\mathbf{Z}; \boldsymbol{\beta}, \sigma^2)) &= \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \sum_{q=1}^Q \sum_{l \geq q}^Q \gamma_{ijql} x_{ijkt} \mathbf{d}_{kt}^\top \boldsymbol{\beta}_{ql} \\ &\quad - \frac{K}{2} \sum_{q=1}^Q \sum_{l \geq q}^Q \log(2\pi\sigma_{ql}^2) - \sum_{k=1}^K \sum_{q=1}^Q \sum_{l \geq q}^Q c_{qlk} \\ &\quad + \sum_{k=1}^K \sum_{q=1}^Q \sum_{l \geq q}^Q \log \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} dr_{qlk}, \end{aligned} \quad (\text{C2})$$

$$h_{qlk} = \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijkt} r_{qlk} + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \log(1 - \pi_{qlkt}) - \frac{r_{qlk}^2}{2\sigma_{ql}^2}. \quad (\text{C3})$$

Next, we state the list of all integrals $I_{qlk1}^c - I_{qlk6}^c$,

$$I_{qlk1}^c = \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} dr_{qlk}, \quad (C4)$$

$$I_{qlk2}^c = \frac{\partial I_{qlk1}^c}{\partial \sigma_{ql}^2} = \frac{1}{2\sigma_{ql}^4} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} r_{qlk}^2 dr_{qlk}, \quad (C5)$$

$$I_{qlk3}^c = \frac{\partial I_{qlk1}^c}{\partial \beta_{ql}} = \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} \pi_{qlkt} dr_{qlk}, \quad (C6)$$

$$I_{qlk4}^c = \frac{\partial I_{qlk2}^c}{\partial \sigma_{ql}^2} = \frac{1}{4\sigma_{ql}^8} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} r_{qlk}^4 dr_{qlk} - \frac{1}{\sigma_{ql}^6} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} r_{qlk}^2 dr_{qlk}, \quad (C7)$$

$$\begin{aligned} I_{qlk5}^c &= \frac{\partial \mathbf{I}_{qlk3}^{c\top}}{\partial \beta_{ql}} = \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right)^2 \sum_{t=1}^T \sum_{t'=1}^T \mathbf{d}_{kt} \mathbf{d}_{kt'}^\top \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} \pi_{qlkt} \pi_{qlkt'} dr_{qlk} \\ &+ \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \mathbf{d}_{kt}^\top \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} \pi_{qlkt} (1 - \pi_{qlkt}) dr_{qlk}, \end{aligned} \quad (C8)$$

$$\begin{aligned} I_{qlk6}^c &= \frac{\partial I_{qlk3}^c}{\partial \sigma_{ql}^2} = \frac{\partial I_{qlk2}^c}{\partial \beta_{ql}} \\ &= \frac{1}{2\sigma_{ql}^4} \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \gamma_{ijql} \right) \sum_{t=1}^T \mathbf{d}_{kt} \int_{-\infty}^{+\infty} e^{h_{qlk} + c_{qlk}} r_{qlk}^2 \pi_{qlkt} dr_{qlk}. \end{aligned} \quad (C9)$$

Equations for the score vector $\mathbf{U}(\beta_{ql}, \sigma_{ql}^2) = (\mathbf{U}(\beta_{ql})^\top, U(\sigma_{ql}^2)^\top)^\top$ are given as

$$\mathbf{U}(\beta_{ql}) = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} x_{ijk} \mathbf{d}_{kt} - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n \sum_{j \neq i}^n \sum_{t=1}^T \gamma_{ijql} \psi_{qlkt} \mathbf{d}_{kt} + \sum_{k=1}^K \frac{I_{qlk3}^c}{I_{qlk1}^c} \quad (C10)$$

$$U(\sigma_{ql}^2) = -\frac{K}{2\sigma_{ql}^2} + \sum_{k=1}^K \frac{I_{qlk2}^c}{I_{qlk1}^c}. \quad (C11)$$

The Fisher information matrix $\mathcal{F}(\beta, \sigma^2)$ is a block diagonal matrix, whose overall dimension is $(P+1)Q(Q+1)/2 \times (P+1)Q(Q+1)/2$. In particular, an individual block component noted as $\mathcal{F}(\beta_{ql}, \sigma_{ql}^2)$ is a $(P+1) \times (P+1)$ matrix, and it can be represented as

$$\mathcal{J}(\beta_{ql}, \sigma_{ql}^2) = \left[\begin{array}{c|c} \mathcal{J}_1(\beta_{ql}) & \mathcal{J}_2(\sigma_{ql}^2, \beta_{ql}) \\ \hline \mathcal{J}_2^\top(\sigma_{ql}^2, \beta_{ql}) & \mathcal{I}_3(\sigma_{ql}^2) \end{array} \right], \quad (\text{C12})$$

such that

$$\mathcal{J}_1(\beta_{ql}) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\mathbf{x}|\mathbf{Z}; \beta, \sigma^2))}{\partial \beta_{ql} \partial \beta_{ql}^\top} = -\sum_{k=1}^K \left[\frac{\mathbf{I}_{qlk5}^c}{\mathbf{I}_{qlk1}^c} - \frac{\mathbf{I}_{qlk3}^c \mathbf{I}_{qlk3}^{c\top}}{(\mathbf{I}_{qlk1}^c)^2} \right], \quad (\text{C13})$$

$$\mathcal{J}_2(\sigma_{ql}^2, \beta_{ql}) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\mathbf{x}|\mathbf{Z}; \beta, \sigma^2))}{\partial \sigma_{ql}^2 \partial \beta_{ql}} = -\sum_{k=1}^K \left[\frac{\mathbf{I}_{qlk6}^c}{\mathbf{I}_{qlk1}^c} - \frac{\mathbf{I}_{qlk3}^c \mathbf{I}_{qlk2}^c}{(\mathbf{I}_{qlk1}^c)^2} \right], \quad (\text{C14})$$

$$\mathcal{I}_3(\sigma_{ql}^2) \equiv -\frac{\partial^2 \mathbb{E}(\log f(\mathbf{x}|\mathbf{Z}; \beta, \sigma^2))}{\partial \sigma_{ql}^2 \partial \sigma_{ql}^2} = -\frac{K}{2\sigma_{ql}^4} - \sum_{k=1}^K \left[\frac{\mathbf{I}_{qlk4}^c}{\mathbf{I}_{qlk1}^c} - \frac{(\mathbf{I}_{qlk2}^c)^2}{(\mathbf{I}_{qlk1}^c)^2} \right]. \quad (\text{C15})$$

Second, we encountered several practical examples in which the function integrate would fail to provide satisfactory estimates of the integrals. This typically occurs in scenarios when the integrand is zero over most of its integration range and the quadrature approximation fails to detect the range of values for which the contribution to the integral is nonzero. We solved this issue by systematically reducing the range of integration. More precisely, this was achieved by first detecting the value of r_{qlk} corresponding to the maximum of h_{qlk} , which is then used as the center of the reduced interval of integration. Then, starting by this value as unique point of integration, we iteratively increased the interval of integration by a value 1 in each direction with the stopping criterion that the values of h_{qlk} at the endpoints of the new interval were inferior to 10^{-10} times the maximum value of h_{qlk} . Noting that h_{qlk} is a concave function, we were sure that all the values outside the reduced interval of integration were also inferior to 10^{-10} times the maximum value of h_{qlk} and hence negligible. To facilitate the convergence of the Newton–Raphson algorithm, we used the same step-halving procedure that has been used in Pavlovic et al. (2020).