Received: 26 February 2019 | Accepted: 4 June 2020

DOI: 10.1111/biom.13319

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Scalable Bayesian matrix normal graphical models for brain functional networks

Suprateek Kundu | **Benjamin B. Risk**

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia

Correspondence
Suprateek Kundu, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322.
Email: suprateek.kundu@emory.edu

Abstract

Recently, there has been an explosive growth in graphical modeling approaches for estimating brain functional networks. In a detailed study, we show that surprisingly, standard graphical modeling approaches for fMRI data may not yield accurate estimates of the brain network due to the inability to suitably account for temporal correlations. We propose a novel Bayesian matrix normal graphical model that jointly models the temporal covariance and the brain network under a separable structure for the covariance to obtain improved estimates. The approach is implemented via an efficient optimization algorithm that computes the *maximum-a-posteriori* network estimates having desirable theoretical properties and which is scalable to high dimensions. The proposed method leads to substantial gains in network estimation accuracy compared to standard brain network modeling approaches as illustrated via extensive simulations. We apply the method to resting state fMRI data from the Human Connectome Project involving a large number of time scans and brain regions, to study the relationships between fluid intelligence and functional connectivity, where it is not computationally feasible to apply existing matrix normal graphical models. Our proposed approach led to the detection of differences in connectivity between high and low fluid intelligence groups, whereas these differences were less pronounced or absent using the graphical lasso.

KEYWORDS

functional connectivity, Human Connectome Project, matrix normal graphical models, precision matrix estimation

1 | INTRODUCTION

Functional connectivity measures how different regions of the brain change their activity together as a functionally coherent unit. Recently, there has been a rapid increase in the development of functional connectivity approaches for resting-state functional magnetic resonance imaging (rsfMRI) in which a subject lies in a scanner looking at cross hairs. Among a variety of proposed methods, the precision matrix estimation approach for brain connectivity was found to be one of the most successful approaches (Smith *et al.*, 2011) in the presence of various issues that often arise in real fMRI studies, given the fact that it can distinguish a true, direct functional connection from an apparent connection between two nodes that can potentially be due to confounding by additional nodes.

The fMRI data recorded from experiments typically contains structured noise that contributes to substantial autocorrelation (Bollmann *et al.*, 2018), arising from physical sources such as scanner drift (eg, slowly varying changes

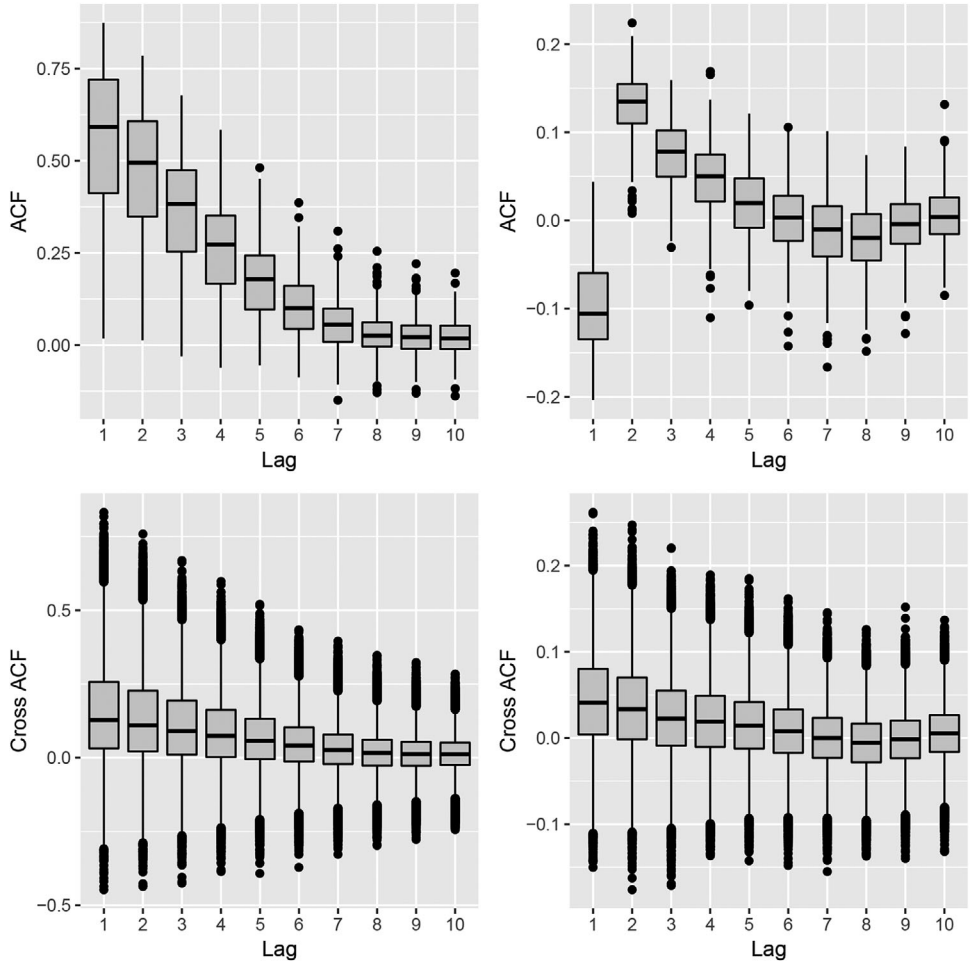

FIGURE 1 Boxplots for autocorrelation and cross-correlation across varying lags for one subject in the Human Connectome Project data. The top panels denote the empirical and prewhitened (using an AR(1) model) autocorrelations (left and right panels, respectively). The bottom panel denotes the corresponding cross-correlations. It is clear that an AR(1) model is unable to completely remove the auto- and cross-correlations from the data

in ambient temperature), physiological factors (eg, breathing and cardiac pulsation), residual movement artifacts, as well as from serial correlations inherent in the blood-oxygen-level dependent signal related to neural activity. Neuroimaging studies typically involve a pipeline of standard preprocessing steps to reduce motion artifacts, which reduce but do not remove serial correlation (Smith *et al.*, 2013). Hence, the preprocessed fMRI data are often temporally correlated within and across brain regions, which may render standard network modeling tools ineffective.

Bayesian Gaussian graphical models (Dobra *et al.*, 2011; Kundu *et al.*, 2019) have been successful in modeling brain networks (Mumford and Ramsey, 2014), and have been recently applied for the joint estimation of multiple networks (Peterson *et al.*, 2015), and for the estimation of dynamic networks (Warnick *et al.*, 2018). Unfortunately, these existing approaches typically do not explicitly account for temporal dependence in fMRI data, and often do not scale to the high-dimensional networks considered in our applications. Although there are some recent asymptotic results for precision matrix estimation using temporally correlated data via the graphical lasso algorithm (Shu and Nan, 2019), the finite sample properties of these approaches are not yet well understood. One possible approach is to minimize the autocorrelations using prewhitening methods (Zhu and Cribben, 2018), which reduces temporal dependence between fMRI measurements, and then fit the graphical models to these prewhitened observations. However, such approaches may not be able to completely remove the temporal correlations as illustrated using the Human Connectome Project (HCP) data in Figure 1. Another alternative is to assume a fixed temporal structure when computing connectivity (Castruccio *et al.*, 2018), which can be restrictive in practical applications with arbitrary temporal dependencies. There is also relevant work on computing connectivity between

channels via partial coherence measures obtained from spectral density matrices for electroencephalogram (EEG) data (Fiecas and Ombao, 2010), but these cannot be directly used for network analysis using fMRI data. One of the important findings of this article is that temporal correlations, when not properly accounted for, may lead to misleading brain network estimation under graphical modeling approaches. Hence, there is a pressing need for approaches that deal with temporal correlations in a systematic manner to improve brain network estimation.

A promising solution may be found by jointly modeling the temporal correlations and the brain network via a separable assumption on the covariance structure. Separable covariance models have been heavily used in the spatial statistics (Genton, 2007), and related literature. Unfortunately these approaches do not focus on estimating conditional dependencies. Moreover, they typically assume prespecified spatiotemporal covariances, or consider low-dimensional multivariate spatiotemporal data, which make them ill-suited for high-dimensional network analysis using graph theoretic approaches. There has also been some development of separable covariance approaches for estimating networks from high-dimensional time series that do not assume any fixed structure on the covariance between nodes in the network (Qiu *et al.*, 2016) but instead assume a prespecified temporal dependence structure that may be restrictive in practical fMRI applications involving long-range dependence (Meisel *et al.*, 2017). Penalized matrix normal graphical models (pMNGM) involving separable covariance structures have also been proposed (Allen and Tibshirani, 2010; Leng and Tang, 2012; Zhou, 2014), and a Bayesian alternative involving hyper inverse-Wishart priors was developed by Wang and West (2009). Although these approaches are attractive in permitting arbitrary structures for the row and column covariances, they are yet to be applied to mainstream brain functional network analysis using high-dimensional fMRI data, to our knowledge.

One potential difficulty with existing MNGM is their computational feasibility for high-dimensional data settings. For example, our motivating HCP application includes rsfMRI scans with 1200 time points, and uses a recently developed multi-modal atlas that contains 360 nodes (over 60 000 edges) (Glasser *et al.*, 2016). As existing pMNGM approaches often involve two tuning parameters for imposing sparsity in both the row and column inverse covariance matrices, they are not suitable for high-dimensional brain network analysis using the HCP data. Moreover, finding optimal choices for two tuning parameters is difficult. Similarly, the Bayesian approaches involving hyper inverse-Wishart priors typically constrain their search to the space of decomposable graphs, which may be restrictive for high-dimensional settings (Kundu *et al.*, 2019).

We propose a novel and computationally scalable Bayesian MNGM (BMNGM) for estimating brain functional networks that bypasses the aforementioned challenges. The proposed approach imposes shrinkage priors on the precision off-diagonals for modeling the brain network and simultaneously models the temporal covariance matrix under an inverse-Wishart prior without relying on sparsity assumptions and additional tuning parameters. In addition to a Markov chain Monte Carlo (MCMC) based implementation of the proposed model, we develop a computationally efficient optimization algorithm to obtain the maximum a posteriori (MAP) estimates. The algorithm yields exact zeros corresponding to absent network edges, and involves only one tuning parameter controlling the network sparsity. The approach scales to a large number of nodes and time points, and the MAP estimator has desirable theoretical properties as the number of nodes and time points grow. We perform extensive simulations to illustrate the practical advantages of the approach. We apply our method to a high-dimensional HCP data application where existing MNGM approaches are not applicable (due to scalability issues) and standard graphical modeling approaches only detect feeble or no network differences related to fluid intelligence changes.

2 | METHODOLOGY

2.1 | Proposed approach

Consider a $T \times v$ matrix of preprocessed fMRI measurements X_i corresponding to T time points and v regions of interests for the ith subject, where the tth row is denoted as $\mathbf{x}_i^{(t)}$ and the jth column is denoted as $\mathbf{x}_{ij}, t = 1, \dots, T,\ j = 1, \dots, v$. In what follows, we will first normalize the columns of the data matrix X (resulting in zero column means and unit variances for each node) before fitting the model, and subsequently perform a postfitting rescaling step to recover the estimated precision matrix (described in Algorithm I in Section 2.3), as in Rothman *et al.* (2008). Consider the Bayesian matrix normal model

$$X_i \sim MN(\mathbf{0}_{T\times v}, \Omega_{v\times v}^{-1} \otimes \mathcal{K}_{T\times T}),\ \Omega \sim \pi(\Omega),\ \mathcal{K} \sim \pi(\mathcal{K}),$$
$$i = 1, \dots, n, \tag{1}$$

where $\mathbf{0}_{T\times v}$ denotes a $T \times v$ matrix of zeros, $\otimes$ denotes the Kronecker product, $MN(\cdot, \cdot)$ implies a matrix normal distribution, Ω denotes the inverse covariance matrix between nodes, and $\mathcal{K}$ encodes temporal correlations. Both Ω and $\mathcal{K}$ are unknown and estimated from the data under

suitable priors $\pi(\Omega)$ and $\pi(\mathcal{K})$ (see Section 2.2). Model (1) is equivalent to assuming a multivariate normal distribution on vec(X) with zero mean and covariance matrix $\Sigma = \Omega^{-1} \otimes \mathcal{K}$, where vec$(X)$ is obtained by stacking the columns of X. As Ω and $\mathcal{K}$ in the likelihood are only identifiable up to a constant, one may optionally rescale $\mathcal{K}$ as in Leng and Tang (2012) to ensure identifiability (see Algorithm I in Supporting Information).

A sparse inverse covariance matrix $\Omega = (\omega_{kl})_{k,l=1}^{v}$ corresponds to a graph that encodes functional connections that is stationary. In other words, the partial correlation between $x_{i,tk}$ and $x_{i,tl}$ is the same as that between $x_{i,t'k}$ and $x_{i,t'l}$, where $t \neq t'$, and $k \neq l$. Denote the graph $\mathcal{G}$ that has the vertex set $\mathcal{V}$ consisting of vertices or nodes $\{1, \ldots, v\}$ and edge set E that contains all edges corresponding to nonzero precision off-diagonals. Edges present in E represent important network connections, while absent edges imply insignificant associations. Further, $\mathcal{K}$ represents the temporal covariances across the scanning session, which are assumed to be constant across all nodes, that is, $\text{corr}(x_{i,tk}, x_{i,t'k}) = \text{corr}(x_{i,tl}, x_{i,t'l})$ (where $l \neq k$). Further, the cross covariance between $x_{i,tk}$ and $x_{i,t'l}$, which represents the associations across nodes at nonzero lags, is given by $\Omega^{-1}(k,l)\mathcal{K}(t,t'), k \neq l, t \neq t'$.

The Kronecker structure on the covariance in (1) is designed to model temporal autocorrelations within a node, and cross-correlations between nodes, while simultaneously estimating Ω. This approach precludes the need to pre-specify a temporal structure and bypasses prewhitening steps, which may otherwise be required to minimize temporal correlations under standard approaches. The Kronecker assumption results in a separable covariance structure that considerably reduces the number of candidate covariance parameters, and provides important benefits such as model parsimony and scalability, which are crucial for a practical implementation in high-dimensional applications. A separable structure implies constant functional connectivity between two nodes across all time points, and conversely, constant temporal correlations across all brain regions. Although the first assumption is commonly used for traditional brain network analyses that do not assume dynamic connectivity, the second constraint implies a global temporal correlation structure. This latter assumption is more flexible than the global autoregressive structures implemented in the popular software SPM, although SPM also includes the option of allowing the coefficients to vary over space (Bollmann *et al.*, 2018). We used numerical studies in Section 4 to examine the separable assumption under the proposed Bayesian model (1), and found it did not significantly affect the ability to accurately estimate the brain network when the true covariance structure is nonseparable. Although the separable assumption may potentially lead to inaccurate temporal correlation estimates, this is tolerated as it is not our primary focus.

2.2 | Prior on the precision matrix and temporal correlations

We propose the Bayesian graphical lasso prior (Wang, 2012) on Ω in (1), which imposes double exponential priors on the precision off-diagonals to shrink the elements corresponding to absent edges toward zero. Further, the temporal correlations in (1) are modeled under an inverse-Wishart prior as follows

$$\pi(\Omega \mid \lambda) \propto \prod_{k=1}^{v} \pi(\omega_{kk}) \prod_{j<k} DE(\omega_{jk}; \lambda) I(\Omega \in M_v^+),$$
$$\mathcal{K} \sim \mathcal{IW}(b, D), \tag{2}$$

where $\pi(\omega_{kk})$ denotes the prior on the diagonal elements in Ω (see Section 2.3), $DE(\lambda)$ represents the double exponential/Laplace prior distribution, with λ being the shrinkage parameter, $\mathcal{IW}$ denotes the inverse-Wishart distribution with b degrees of freedom and scale matrix D, and M_v^+ refers to the space of all $v \times v$ positive definite matrices. We note that by varying λ, one can still obtain a series of networks corresponding to varying sparsity levels, with a greater value of λ resulting in more sparsity and vice versa. We denote the proposed approach in (1) and (2) as the BMNGM.

2.3 | Computational algorithm

The posterior distribution for the parameters in our model (1) and (2) is given as $\pi(\Omega, \mathcal{K} \mid -) \propto L(X_1, \ldots, X_n \mid \Omega, \mathcal{K})\pi(\Omega \mid \lambda)\pi(\mathcal{K})$. For the MCMC implementation, we assume that the diagonal elements of Ω are independently distributed as Exp$(\lambda/2)$. The posterior computation can proceed via fully Gibbs steps that performs columnwise updates for the elements in Ω using a normal posterior under a scale mixture representation of the double exponential distribution (Wang, 2012), followed by updating $\mathcal{K}$ using an inverse-Wishart posterior. The shrinkage parameter λ is treated as random and also updated under a Gamma prior. The MCMC details along with the thresholding procedure to obtain point estimates of the precision matrix is detailed in Web Appendix C of the Supporting Information.

For MAP estimation λ is treated as a tuning parameter. The negative log-posterior is

$$Q(\Omega,\mathcal{K}) = \Big\{ -\frac{nT}{2}\log(\det(\Omega)) + \frac{n\upsilon}{2}\log(\det(\mathcal{K})) + \frac{1}{2}Tr\{\mathcal{K}^{-1}\sum_{i=1}^{n}(X_i\Omega X_i^T)\}\Big\} - \log(\pi(\Omega;\lambda)) + \frac{b+T+1}{2}\log(\det(\mathcal{K})) + \frac{1}{2}Tr(D\mathcal{K}^{-1}) + C, \tag{3}$$

where $Tr(\cdot)$ and $\det(\cdot)$ denote the trace and determinant, C denotes a constant not involving model parameters, and the term $-\log(\pi(\Omega;\lambda)) \propto \lambda\sum_{j\le k}|\omega_{jk}|$ or $\lambda\sum_{j<k}|\omega_{jk}|$ depending on whether the prior on the diagonal elements is exponential or uninformative (ie, $\pi(\omega_{kk}) \propto 1$). Although the first choice results in a graphical lasso algorithm (Friedman *et al.*, 2008), the second one resembles the related method in Rothman *et al.* (2008), which is what we will use for MAP estimation. Even under the uniform prior $\pi(\omega_{kk}) \propto 1$, one obtains a proper posterior as shown in Lemma 1 (see Web Appendix A of Supporting Information for proof).

Lemma 1. *Given model (1), the prior $\pi(\Omega \mid \lambda)$ in (2) with $\pi(\omega_{kk}) \propto 1, k = 1,\dots,\upsilon$, results in a proper posterior distribution $\pi(\Omega \mid \lambda, \mathcal{K}, X_1,\dots,X_n)$.*

We propose a coordinate descent algorithm for obtaining the MAP estimate for parameters corresponding to (3), which iteratively updates Ω and $\mathcal{K}$ till convergence. Algorithm I in Web Appendix C of Supporting Information provides full details. The proposed approach can be fit for a series of λ values resulting in varying sparsity levels, and the optimal graph can be chosen corresponding to the λ value minimizing the Bayesian information criteria (BIC). Alternatively, the value of λ is chosen to give a desired network sparsity in brain imaging applications. The total cost of the algorithm at an iteration with κ^* nonzero elements in Ω is between $O(\upsilon\kappa^*) + O(T^{2.37})$ and $O(\upsilon\kappa^*) + O(T^3)$. We note that the computation time is smaller for larger values of the penalty parameter that results in more sparse precision matrices. Although there is no guarantee that the algorithm converges to the global minimum, it is possible to show that the algorithm converges to a local stationary point using a similar argument as in Proposition 2 of Allen and Tibshirani (2010). See Lemma 2 (proof very similar to the proof of Proposition 2 in Allen and Tibshirani (2010)). Moreover, we illustrate in Section 3 that the proposed MAP estimator is able to asymptotically recover the true precision matrix under certain commonly used regularity conditions.

Lemma 2. *Iterative block coordinate-wise maximization of $Q(\Omega,\mathcal{K})$ with respect to Ω and $\mathcal{K}$ converges to a stationary point of $Q(\Omega,\mathcal{K})$.*

3 | THEORETICAL PROPERTIES

In our theoretical treatment, we show the consistency of the MAP estimator for Ω as the number of brain voxels and time scans increase, while the sample size is fixed (without loss of generality, fix $n = 1$). Unlike some previous work on MNGM (Zhou, 2014) who showed consistency in estimating both the row and column covariances, we are only interested in the accurate estimation of the precision matrix that is of practical interest in brain network studies, and we consider $\mathcal{K}$ as a nuisance parameter. Moreover, our theoretical results focus on a broad class of true models (not necessarily MNGM) that capture several practical settings (see below). In particular, we assume that the true model is such that $\mathbf{x}^{(t)}, t = 1,\dots,T$, follow the same distribution with $\mathrm{Cov}(X_t) = \Sigma_0 = \Omega_0^{-1}$ and corresponding correlation matrix R_0, where Ω_0 belongs to the class of well-behaved matrices as elaborated in Assumption (**C1**) in the sequel. We assume that the true data generating mechanism follows

$$\mathrm{vec}(X) = H\mathbf{e},\ E(\mathbf{e}) = \mathbf{0},\ \mathrm{Cov}(\mathbf{e}) = I_m, \tag{4}$$

where I_m is an $m \times m$ identity matrix, $H = (h_{ij})$ is a $\upsilon T \times m$ real deterministic matrix, and $\mathbf{e} = (e_1,\dots,e_m)$ is a random vector with e_j and $e_{j'}$ independent ($j \neq j'$). Moreover, the true covariance matrix for $\mathrm{vec}(X)$ is given by $\mathrm{Cov}(H\mathbf{e}) = H\mathrm{Cov}(\mathbf{e})H^T = HH^T$. We assume that the true covariance parameters in Σ_0 are bounded even when the number of columns in H (ie, m), goes to ∞. The class of models defined by (4) covers multivariate fractional Gaussian noise and vector autoregressive models, as discussed in Shu and Nan (2019).

Denote $\Lambda(M)$ as the eigen-values of M, and define the spectral, Frobenius, and matrix L_1 norms as $||M||_2 = \sqrt{\Lambda_{\max}(M'M)}$, $||M||_F = \sqrt{\sum_{i,j}M_{ij}^2}$, and $||M||_1 = \max_j\sum_k M_{kj}$, respectively. The entry-wise L_1 norm is $|M|_1 = \sum_{kl}|M|_{kl}$ and the corresponding off-diagonal version is $|M|_{1,off} = \sum_{k\neq l}|M|_{kl}$, and finally, the entry-wise L_∞ norm $M_\infty = \max_{k,l}|M_{kl}|$. Let $R_{[k]}$ be the $T \times T$ correlation matrix for the kth time series, $k = 1,\dots,\upsilon$. Further, let S_0 denote the true edge set, that is, the set of non-zero entries for Ω_0, S_0^c denote the complement of this set, and define d^* to be the maximum number of nonzeros per row in Ω_0. Define $R_0^{\otimes 2} = R_0 \otimes R_0$ as the Kronecker product between R_0 with itself, and let $R_{0,S,S'}^{\otimes 2}$ denote a submatrix of $R_0^{\otimes 2}$ with rows and columns

corresponding to $S, S' \subseteq \{1, \ldots, \upsilon\}$. Define $\kappa_R = ||R||_1$ and $\kappa_{R_0^{\otimes 2}} = ||(R_0^{\otimes 2})^{-1}||_1$. Assumptions (**C1**)-(**C4**) are described (see Web Appendix A of Supporting Information for further discussions).

(**C1**): The class of true precision matrices satisfy $\Omega_0 = \Sigma_0^{-1} \in \mathcal{G}_2(s_\upsilon, \nu_0) := \{\Omega_0 : \sum_{1 \le k < k' \le \upsilon} 1(\omega_{0,kl} \neq 0) \le s_\upsilon, 0 < \nu_0^{-1} \le \Lambda_{\min}(\Omega) \le \Lambda_{\max}(\Omega) \le \nu_0\}$, where s_υ determines the sparsity of the graph and can depend on ν_0.

(**C2**): $\max_{1 \le k \le \upsilon} \frac{1}{\upsilon} ||R_{[k]}||_F^2 \le g_F$, and $\max_{1 \le k \le \upsilon} ||R_{[k]}||_2 \le g_2$, with $1 \le g_F \le g_2 \le T$.

(**C3**): $\max_{e \in S_0^c} |R^{\otimes 2}_{0,eS_0} (R^{\otimes 2}_{0,S_0S_0})^{-1}|_1 \le 1 - \beta$, for some $\beta \in (0, 1]$.

(**C4**): Consider a sequence of priors on the temporal correlation matrix $\pi(\mathcal{K})$ such that $b \to \infty, b - \upsilon > 0$, and $D/b \to I_T$ as $T, \upsilon \to \infty$.

Define $u_1 = \max\{\log(\upsilon) g_2/T, \sqrt{\log(\upsilon) g_F/T}\}$ and $r = (0.5 + 2.5(1 + 8/\beta)\kappa_{R_0^{\otimes 2}}) M u_1 \nu_0$. Theorem 1 is presented (proof provided in Web Appendix A of Supporting Information).

Theorem 1. *Suppose $X_{T \times \upsilon}$ is generated from (4), with all e_j satisfying $(E|e_j|^k)^{1/k} \le K k^{1/2}$ where K is a positive constant, for all $k \ge 1$. Further, let assumptions (**C1**)-(**C4**) hold, for sufficiently large constant M depending on ν_0 and K, and $\lambda = 8 M u_1/\beta \le (6(1+\beta)/8) d^* \max\{\kappa_{R_0} \kappa_{R_0^{\otimes 2}}, \kappa^3_{R_0} \kappa^3_{R_0^{\otimes 2}}\})^{-1}$, and $u_1 = o(\min\{1, ((1 + 8/\beta)\kappa_{R_0^{\otimes 2}})^{-1}\})$. Then as $\upsilon, T \to \infty$, with probability tending to one, the following results hold:*

(i) $|\hat{\Omega}_\lambda - \Omega_0|_\infty \le r$, $||\hat{\Omega}_\lambda - \Omega_0||_2 \le r \min\{d^*, \upsilon + s_\upsilon\}$, *and* $\frac{1}{\sqrt{\upsilon}} ||\hat{\Omega}_\lambda - \Omega_0||_F \le r\sqrt{1 + s_\upsilon/\upsilon}$.

(ii) For all $(k, l) \in S_0^c$, $\hat{\omega}_{kl} = 0$, and under the additional assumption that $|\omega_{0,kl}| > r$ for all $(k, l) \in S_0$, we have $sign(\hat{\omega}^*_{kl}) = sign(\omega_{0,kl})$.

4 | SIMULATION STUDY

4.1 | Set-up

We performed a variety of simulation studies and compared the performance of the proposed approach with existing Gaussian graphical modeling approaches that (a) fit the network using temporally correlated data; (b) use prewhitening to minimize temporal correlations and then fit the network; and (c) jointly model the brain network and temporal correlations. All results in this section under BMNGM correspond to the MAP estimator. We compare our results with graphical lasso for both prewhitened data (Glasso-AR), and data without prewhitening (Glasso). The prewhitening was performed using the "auto.arima" function in R ("forecast" package), which automatically selects the autoregressive lag. We also compared with the pMNGM that assumes a similar separable covariance structure as BMNGM but imposed sparsity in both Ω and $\mathcal{K}^{-1}$ via two distinct penalty parameters. The pMNGM is fit exactly as Algorithm I for BMNGM, but uses Glasso to estimate $\mathcal{K}^{-1}$. We also compare with the BMNGM approach that has a fixed temporal dependence structure $\mathcal{K} = (n\upsilon)^{-1} \sum_{i=1}^n \sum_{j=1}^\upsilon (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{ij})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{ij})^T$, which is denoted as empirical MNGM (eMNGM). We note that Glasso was implemented via the "QUIC" package in R (Hsieh *et al.*, 2011).

We generate data involving $n = 20$ subjects under multiple scenarios. Scenario I generated data from the proposed model with two types of networks and temporal correlation structures. Scenario II incorporated polynomial decay temporal correlations such that $\text{Cov}(\mathbf{x}_{it}, \mathbf{x}_{i,t+j}) = (j+1)^{-\alpha} \Omega_0^{-1}$. Hence, the covariance has a separable structure, but the time series for each node is generated with polynomial temporal correlations, similar to Shu and Nan (2019). For Scenarios I and II, we generate the data for $\upsilon = 100, 200$, and $T = 200, 500$. Scenario III generates data under a nonseparable covariance structure. The full details are provided in Web Appendix D of Supporting Information. To assess numerical performance, we compute the area under the curve (AUC) under the receiver operating characteristic curves, which is a measure of the estimated sensitivity versus specificity over different network densities. Matthews Correlation Coefficient (MCC) (Wang, 2012), along with the L_1 error for precision matrix estimation were also used to measure network estimation accuracy. For BMNGM, eMNGM, and Glasso, these measures are reported under the point estimates of the network that are obtained by minimizing the BIC as in Yuan and Lin (2007). For pMNGM, it may not be straightforward to obtain the optimal choices of the two tuning parameters required to impose sparsity in Ω (denoted by λ_1) and $\mathcal{K}^{-1}$ (denoted by λ_2). Hence, we report the highest AUC across a series of λ_2 values. For the corresponding λ_2 value, we report the lowest L_1 error across the grid of λ_1 values and the associated MCC, reflecting a best case scenario under pMNGM.

4.2 | Results

The results in Table 1 suggest that the proposed approach has a comparable or improved AUC and L_1 error compared to alternate approaches under Scenario I. Moreover, the MCC value for BMNGM is higher compared to the other methods for the majority of cases. Except two cases, the eMNGM approach has a lower or comparable MCC and AUC compared to BMNGM. Additional simulation studies with Erdos-Renyi networks with $\upsilon = 200, T = 500$, for

TABLE 1 Numerical results for data generated under a true separable covariance in Scenario I

	Scale-free network, AR(1) temporal covariance											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.92	**0.84**	0.01	0.93	**0.84**	**0.01**	**0.89**	0.74	0.002	0.91	**0.81**	0.002
eMNGM	0.91	**0.84**	0.01	0.91	**0.84**	**0.01**	0.83	0.64	0.002	0.80	0.42	0.002
pMNGM	0.90	0.18	0.02	–	–	–	0.58	0.20	0.006	-	-	-
GLasso - AR	0.91	0.41	0.02	0.93	0.48	0.03	0.82	0.69	0.002	0.89	0.64	0.003
GLasso	0.91	0.41	0.02	0.92	0.42	0.03	**0.88**	0.73	0.002	0.91	0.71	0.003
	Scale-free network, squared exponential temporal covariance											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.91	**0.81**	0.01	0.92	**0.77**	**0.01**	0.89	0.72	**0.001**	**0.91**	0.72	0.002
eMNGM	0.91	0.64	0.01	0.91	0.64	**0.01**	0.81	0.34	**0.001**	0.79	0.59	0.002
pMNGM	0.90	0.14	0.02	–	–	–	0.89	0.16	0.006	-	-	-
GLasso - AR	0.91	0.55	0.02	0.92	0.59	0.03	0.87	**0.74**	0.002	0.88	**0.83**	0.002
GLasso	0.90	0.60	0.02	0.92	0.61	0.04	0.88	**0.75**	0.002	**0.90**	**0.83**	0.002
	Small world network, AR(1) temporal covariance											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.93	**0.85**	**0.003**	0.93	0.81	0.02	0.90	0.75	0.006	0.93	**0.82**	**0.006**
eMNGM	0.92	**0.84**	0.006	0.94	**0.84**	0.02	0.88	0.75	0.006	0.91	0.76	**0.005**
pMNGM	0.90	0.40	0.019	–	–	–	0.78	0.45	0.009	-	-	-
GLasso - AR	0.92	0.34	**0.003**	0.94	0.41	0.07	0.89	0.75	0.006	0.92	0.45	0.008
GLasso	0.91	0.37	0.01	0.94	0.37	0.03	0.89	0.75	0.006	0.92	0.43	0.008
	Small world network, squared exponential temporal covariance											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.92	**0.82**	0.004	0.93	**0.80**	0.02	0.88	**0.73**	**0.005**	0.92	0.72	0.005
eMNGM	0.92	0.65	0.005	0.92	0.66	0.02	0.88	0.53	**0.005**	0.91	**0.81**	0.006
pMNGM	0.90	0.41	0.019	–	–	–	0.79	0.42	0.009	-	-	-
GLasso - AR	0.91	0.40	0.004	0.93	0.37	0.03	0.88	0.44	0.007	0.92	0.74	0.007
GLasso	0.91	0.46	0.01	0.92	0.49	0.03	0.88	0.50	0.007	0.91	0.79	0.007

Note. Glasso-AR refers to the graphical lasso approach using prewhitened observations under an AR model, and other methods are described in Section 4. We do not report the results under the penalized MNGM (pMNGM) for $T = 500$ as the algorithm failed to converge. Bolded numbers imply a significantly better performance compared to at least two other methods.

Scenario I (results presented in Table 1 in Supporting Information) suggested that BMNGM has similar AUC but higher MCC values as compared to the eMNGM. This implies that the proposed approach is more suitable for a more diverse variety of true networks compared to eMNGM. The graphical lasso approaches with and without prewhitening perform worse than BMNGM in general, with the exception of scale-free networks for $v = 200$ where they have higher MCC values under the squared exponential covariance. However, even in these cases the BMNGM has a comparable or higher AUC and comparable or lower L_1 error.

Under Scenario II (see results in Table 2), the BMNGM approach has comparable or significantly higher AUC compared to other methods. Similarly, the MCC values under BMNGM is higher for most cases. For the two cases where the graphical lasso with prewhitening has higher MCC, it is seen that the BMNGM has a comparable L_1 error and a comparable or higher AUC. In general, the L_1 error under BMNGM is almost always comparable or lower relative to the other methods, while the L_1 error for graphical lasso without prewhitening is often the highest. On the other hand, the pMNGM approach failed to converge for all cases in Scenario II, indicating its inadequacy for certain cases when the true correlations are long range. The performance under eMNGM also seems to suffer in Scenario II for $T = 500$, which may be due to nonnegligible long range correlations that could be difficult to estimate empirically. Under Scenario III, the proposed BMNGM approach and the related eMNGM method consistently

TABLE 2 Numerical results for data generated under a vector autoregressive model with a true separable covariance structure (Scenario II)

	Small world network											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.94	**0.76**	0.002	**0.96**	0.20	**0.002**	**0.86**	**0.50**	0.003	**0.90**	0.63	**0.003**
eMNGM	0.96	**0.76**	0.003	0.67	0.11	0.007	**0.87**	0.19	0.004	0.75	0.17	0.007
GLasso-AR	0.94	0.53	0.002	**0.94**	**0.52**	**0.002**	0.81	0.46	0.002	**0.90**	**0.83**	**0.002**
GLasso	0.81	0.07	0.041	0.80	0.27	0.037	0.72	0.25	0.003	0.68	0.06	0.03
	Scale-free network											
	$v = 100, T = 200$			$v = 100, T = 500$			$v = 200, T = 200$			$v = 200, T = 500$		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.96	**0.78**	0.002	**0.96**	**0.63**	**0.003**	**0.87**	**0.52**	0.003	**0.69**	**0.20**	**0.002**
eMNGM	0.97	0.75	0.004	0.75	0.17	0.007	**0.88**	0.17	**0.001**	0.67	0.11	0.007
GLasso - AR	0.94	0.54	0.002	**0.94**	0.54	**0.002**	0.84	0.35	0.002	0.57	0.17	**0.001**
GLasso	0.86	0.09	0.035	0.84	0.06	0.032	0.76	0.30	0.002	0.35	0.02	0.037

Note. Glasso-AR refers to the graphical lasso approach using prewhitened observations under an AR model, and the other methods are described in Section 4. We do not report the results under the penalized MNGM (pMNGM) as the algorithm failed to converge or encountered numerical instability. Bolded numbers imply a significantly better performance compared to at least two other methods.

TABLE 3 Results for data generated under a nonseparable covariance structure (Scenario III) with $v = 40, T = 100$

	Erdos-Renyi			**Small world**			**Scale free**		
Methods	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**	**AUC**	**MCC**	**L1 error**
BMNGM	0.91	**0.79**	**0.018**	0.88	0.74	**0.040**	0.89	**0.71**	**0.033**
eMNGM	0.91	**0.79**	**0.018**	0.88	0.74	**0.039**	0.88	**0.72**	**0.032**
pMNGM	0.88	0.30	0.038	0.86	0.19	0.058	0.84	0.16	0.052
GLasso - AR	0.91	0.67	0.022	0.88	0.73	0.044	0.89	0.69	0.038
GLasso	0.90	0.65	0.022	0.88	0.74	0.044	0.89	0.69	0.038

[a]*Note*. Glasso-AR refers to the graphical lasso approach using prewhitened observations under a AR model, and the other methods are described in Section 4. Bolded numbers imply a significantly better performance compared to at least two other methods.

has a significantly lower L_1 error and significantly higher MCC. Given the fact that the separability assumption is violated in Scenario III, the results under BMNGM for this case is quite encouraging, especially in contrast to the graphical lasso based approaches.

Moreover, the consistently poor performance under pMNGM (even under one of the best case scenarios), coupled with the lack of scalability, illustrate the challenges of implementing this approach for brain network analysis using high dimensional fMRI data. We conjecture that the poor performance under pMNGM may be attributed to approximating $\mathcal{K}^{-1}$ by a sparse matrix that may not be realistic. On the other hand, the BMNGM approach only takes a few (often 40 or less) iterations to converge for the simulation scenarios considered, and in general, performs well for experiments with $T > 100$ (that is almost always the case in practical fMRI studies) and nonnegligible sample size (> 10). Additional simulation results for varying number T are reported in Table 2 of Supporting Information.

5 | ANALYSIS OF HUMAN CONNECTOME PROJECT DATA

5.1 | Description

We examined the relationship between fluid intelligence (hereafter gF) and functional connectivity using the HCP data. gF measures a person's ability to solve problems without past knowledge and was measured using the Penn Matrix Reasoning Task A (PMAT24_A_CR). The impact of functional connectivity on gF is an area of active research without a consensus in the neuroscience community. For example, the MegaTrawl web interface (April 2016 "HCP820-MegaTrawl" release) allows a user to query the relationship between HCP behavioral variables and functional connectivity. In univariate regression of gF versus functional connectivity for each node of a 200-node network, no edges survived multiplicity corrections. This motivates the development of statistical methodology based on partial correlations accounting for

TABLE 4 Difference in graph theoretic measures between high and low fluid intelligence groups

	BMNGM-MAP		BMNGM-MCMC	GLasso	
Measure	**Difference**	***P*-value**	**Difference**	**Difference**	***P*-value**
Degree	0.0122	0.0110	0.008	−0.0015	0.2608
Clustering coefficient	**−0.0210**	**0.0002**	−0.0150	−0.0083	0.0107
Characteristic path length	**−0.0461**	**0.0062**	−0.0378	−0.0018	0.6384
Local efficiency	**−0.0075**	**0.0003**	−0.0044	**−0.0055**	**0.0067**
Global efficiency	0.0118	0.0075	0.0138	−0.0002	0.8403
Modularity	−0.0308	0.0164	−0.0408	−0.0052	0.3317
Assortativity	0.0131	0.3166	0.0191	0.0026	0.7995

[a]*Note*. Differences in bold are significant after Bonferroni correction.

temporal correlations. We used the multimodal parcellation atlas (Glasser *et al.*, 2016) involving 360 regions that were grouped into six consensus communities: (a) visual; (b) somatomotor; (c) dorsal salience; (d) ventral salience; (e) default mode; and (f) central executive (Akaki and Abdallah, 2018) for our analysis. The scan for each individual was 14:33 minutes with TR = 0.72, resulting in 1200 time points for each voxel. The details for fMRI data preprocessing can be found in Web Appendix E of Supporting Information.

To examine whether partial correlations are related to differences in fluid intelligence, we dichotomized fluid intelligence by subsetting to subjects with low or high gF, and then estimated the partial correlations for each group. In our initial analysis, we found that the relationship between gF and partial correlations was confounded by gender. To avoid possible confounding, we restricted the analysis to females. We used the subset of subjects having fluid intelligence scores in the lower (gF $\leq$ 13) and upper (gF $\geq$ 21) quartiles to form the low and high fluid intelligence groups with 149 and 137 individuals, respectively. We examined network differences in: (a) global degree, clustering coefficient, characteristic path length, average local efficiency, efficiency, modularity, and assortativity; (b) nodal measures including degree, clustering coefficient, characteristic path length, and local efficiency; and (c) edge-level partial correlations. Graph theoretic measures were calculated using the Brain Connectivity Matlab toolbox (Rubinov and Sporns, 2010).

Permutation testing was used to infer significant network differences. We permuted the group memberships for a random subset of individuals and estimated precision matrices for both permuted groups using a penalty selected to yield approximately 15% nonzero edges, which resulted in $\lambda = 0.1$. This was repeated 10 000 times to obtain a permutation distribution for differences in edge strengths and network summary measures between groups, which was used for a permutation test. Specifically, the permutation *P*-value was calculated as the proportion of times the absolute value of the difference in the original sample was less than the absolute value of the difference in a permuted sample. MAP estimation takes approximately 20 minutes for each data set on a 3 GHz processor, which was implemented on a cluster with 40 processors. Additionally, we conducted the same analysis but using partial correlations estimated using GLasso ignoring temporal correlation. For fair comparison, the penalty was chosen as $\lambda = 0.03$ to result in approximately 15% network density.

5.2 | Results

There were notable differences in global connectivity measures between the high and low fluid intelligence groups when using the MAP estimator for the BMNGM, while the differences were absent or less significant under GLasso (Table 4). Under BMNGM, degree and global efficiency were higher in high gF ($P = 0.01$ and $P = 0.0075$, respectively) and characteristic path length was lower ($P = 0.0062$), which indicates a greater number of connections and faster information transfer in the high gF group. Clustering coefficient and local efficiency were significantly lower in the high gF group ($P < 0.007$), and modularity was also lower ($P = 0.02$). This implies that the connections in the brain for the high gF group were less clustered and more evenly distributed than the connections in the network for the low gF group. The MCMC-based estimates for BMNGM were similar to the MAP estimates highlighting the robustness of the proposed approach. As the MCMC samples for the precision matrix are only able to provide a point estimate for the network, it is not straightforward to test for significant differences of network summary measures under the MCMC implementation. For Glasso, these measures were in the same direction as in BMNGM but the only significant difference was with respect to local efficiency, which suggests a reduced ability to detect significant differences in global network features compared to BMNGM.

At the node level, we examined the distribution of the *P*-values (by community) for each of the four nodal

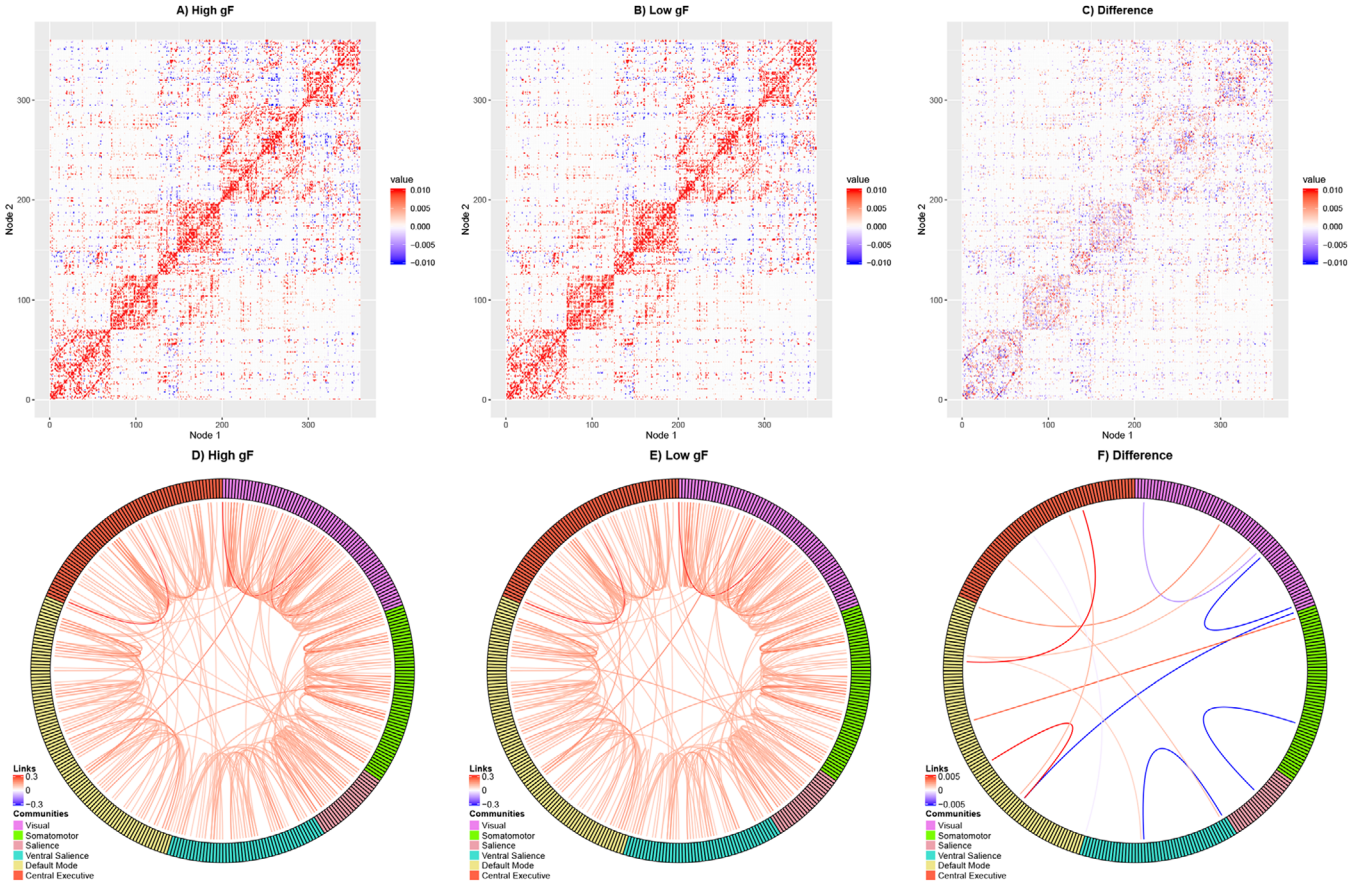

FIGURE 2 This figure appears in color in the electronic version of this article, and any mention of color refers to that version. MAP estimates for partial correlations in high gF (A), low gF (B), the difference between the two (C), and circle plots for the high gF (D), low gF (E), and difference in edge strengths (corresponding to $P < 0.0001$), where red indicates edge strength in gf high > gf low and blue indicates gf high < gf low (F). In A-C, the nodes are ordered according to hierarchical clustering with six functional communities in (Akiki and Abdallah, 2019). Indices 1-70 – visual; indices 71-125 – somatomotor; indices 126-148 – dorsal salience; indices 149-197 – ventral salience; indices 198-293 – default mode; indices 294-360 – central executive

network measures. A higher density near zero is evidence that *P*-values are smaller than expected under the null, and we see there is evidence of differences in nodal network measures for BMNGM while less for GLasso (see Figure 1 in Supporting Information). In the BMNGM, the *P*-values cluster near zero for the central executive and default mode communities for node degree and characteristic path length. The central executive network includes portions of the frontoparietal cortex, which is thought to play an important role in fluid intelligence (Finn *et al.*, 2015). With respect to degree, there were 17 nodes with $P < 0.01$ in BMNGM versus 6 in GLasso; clustering coefficient: 17 versus 8; characteristic path length: 39 versus 7; and local efficiency: 17 versus 6. In BMNGM, nodes R_a10p (central executive) and R_OP2-3 (somatomotor) were notable for having very small *P*-values for degree, clustering coefficient, characteristic path length, and local efficiency (R_a10p and R_OP2-3 correspond to indices 311 and 87 in Figure 2). This Figure appears in color in the electronic version of this article, and any mention of color refers to that version. Note that using a Bonferonni correction of 0.05/360 and 10 000 permutations, a graph measure differs between high and low groups if at most one permutation has a statistic greater than it (ie, $P \leq 0.0001$).

For node R_a10p, the differences between the high and low gF groups for degree, clustering coefficient, characteristic path length, and local efficiency were 0.05 ($P = 0$), −0.17 ($P = 0.0001$), −0.19 ($P = 0.0021$), and −0.07 ($P = 0.0009$). Hence, our analysis discovers a pivotal role for this node in fluid intelligence. The node is located in the frontal lobe and associated with the central executive module. We can visualize the role of node R_a10p and its possible connections by calculating partial correlation differences between the high and low gF groups, and then plotting this difference with the remaining nodes (see Figure 2 in Supporting Information). The differences

tend to concentrate in the frontal lobe, which plays an important role in decision-making, memory formation, and other cognitive functions. Unfortunately, none of the node level differences survived multiplicity adjustment under graphical lasso.

At the edge level, partial correlations tended to be nonzero in contralateral regions—see Figure 2. Further, edges were more common within the functional communities than between modules. Partial correlation heatmaps for the 25th, 50th, and 75th quantiles for both groups under MCMC are also reported in Figure 3 in Supporting Information, which show similar patterns. There were 14 edges with very strong associations (zero P-values) based on 10 000 permutations (see Figure 2F and Table 3 in Supporting Information). Ten of these edges were inter-community edges. Nine edges included a node in the default mode, and four included a node in the central executive module. The results of our HCP analysis signify important whole brain network differences with respect to behavioral and cognitive measures in high-dimensional HCP data, and addresses an important gap in literature.

6 | DISCUSSION

We have proposed a novel BMNGM approach for the estimation of brain networks while accounting for temporal dependence. In the likelihood (3), the inverse of the temporal covariances, $\mathcal{K}^{-1}$ weights the partial covariances via the term $Tr\{\mathcal{K}^{-1}\sum_{i=1}^{n} X_i \Omega X_i^T\}$. We conjecture this leads to improvements in partial correlation estimates. Our approach is designed for group level analyses with at least a small number of subjects, but may not perform well for exceedingly small sample sizes ($n < 10$). We note that it is possible to extend the proposed approach to mixtures of BMNGM in order to tackle between subject heterogeneity, which would discover subgroups of individuals with distinct networks. Extensions can also be made for computing functional connectivity guided by anatomical knowledge (Kang *et al.*, 2017; Higgins *et al.*, 2018) that account for temporal dependence in fMRI, by including edge-specific shrinkage parameters that are modeled in terms of brain structural connectivity. We note one could potentially consider an alternate Bayesian analysis that estimates the precision matrices for both high and low intelligence groups jointly by pooling information across networks and under a common choice of λ, and then thresholds the precision matrices at each MCMC iteration to obtain an approximate sample of networks for both groups. Such a joint analysis could result in greater accuracy by sharing information across groups and provide a more elegant mechanism to estimate and infer network differences at multiple scales using the MCMC samples of the networks. However, more work is needed for developing scalable models for jointly estimating multiple networks that also account for temporal dependence in fMRI. We plan to investigate these issues in future work.

ACKNOWLEDGMENTS

Data were provided (in part) by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657). The authors would like to thank Ixavier Higgins and Xin Ma for providing helpful network and plotting scripts.

DATA SHARING STATEMENT

The data that support the findings in this paper are openly available at http://www.humanconnectomeproject.org/data/.

ORCID

Suprateek Kundu https://orcid.org/0000-0002-1767-4875

REFERENCES

Akiki, T.J. and Abdallah, C.G. (2019) Determining the hierarchical architecture of human brain using subject-level clustering of functional networks. *Scientific Reports*, 9(1), 1–15.

Allen, G.I. and Tibshirani, R. (2010) Transposable regularized covariance models with an application to missing data imputation. *The Annals of Applied Statistics*, 4(2), 764–790.

Bollmann, S., Puckett, A.M., Cunnington, R. and Barth, M. (2018) Serial correlations in single-subject fMRI with sub-second TR. *NeuroImage*, 166, 152–166.

Castruccio, S., Ombao, H. and Genton, M.G. (2018) A scalable multiresolution spatiotemporal model for brain activation and connectivity in fMRI data. *Biometrics*, 74, 823–833.

Dobra, A., Lenkoski, A. and Rodriguez, A. (2011) Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106(496), 1418–1433

Fiecas, M. and Ombao, H. (2010) The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Annals of Applied Statistics*, 5(2A), 1102–1125.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, D. and Constable, R.T. (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), 1664–1671.

Friedman, J., Hastie, T. and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics*, 9(3), 432–441.

Genton, M.G. (2007) Separable approximations of space–time covariance matrices. *Environmetrics*, 18(7), 681–695.

Higgins, I.A., Kundu, S. and Guo, Y. (2018) Integrative Bayesian analysis of brain functional networks incorporating anatomical knowledge. *Neuroimage*, 181, 263–278.

Hsieh, C., Sustik, M.A., Dhillon, I.S. and Ravikumar, P. (2011) Sparse inverse covariance matrix estimation using quadratic

approximation. *Advances in Neural Information Processing Systems*, 24, 2330–2338.

Kang, H., Ombao, H., Fonnesbeck, C., Ding, Z. and Morgan, V.L. (2017) A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. *Brain Connect*, 7(4), 219–227.

Kundu, S., Mallick, B.K. and Baladandayuthapani, V. (2019) Efficient Bayesian regularization for graphical model selection. *Bayesian Analysis*, 14(2), 449–476.

Leng, C. and Tang, C.Y. (2012) Sparse matrix graphical models. *Journal of the American Statistical Association*, 107(499), 1187–1200.

Meisel, C., Bailey, K., Achermann, P. and Plenz, D. (2017) Decline of long-range temporal correlations in the human brain during sustained wakefulness. *Scientific Reports*, 7(1), 11825.

Mumford, J.A. and Ramsey, J.D. (2014) Bayesian networks for fMRI: a primer. *Neuroimage*, 86, 573–582.

Peterson, C., Stingo, F.C. and Vannucci, M. (2015) Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*, 110(509), 159–174.

Qiu, H., Han, F., Liu, H. and Caffo, B. (2016) Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2), 487–504.

Rothman, A.J., Bickel, P.J., Levina, E. and Zhu, J. (2008) Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2, 494–515.

Rubinov, M. and Sporns, O. (2010) Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52, 1059–1069.

Shu, H. and Nan, B. (2019) Estimation of large covariance and precision matrices from temporally dependent observations. *Annals of Statistics*, 47, 3, 1321–1350.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., et. al., (2013) Resting-state fMRI in the human connectome project. *Neuroimage*, 80, 144–168.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D. and Woolrich, M.W. (2011) Network modelling methods for FMRI. *Neuroimage*, 54(2), 875–891.

Wang, H. (2012) Bayesian Graphical Lasso models and efficient posterior computation. *Bayesian Analysis*, 7(4), 867–886.

Wang, H. and West, M. (2009) Bayesian analysis of matrix normal graphical models. *Biometrika*, 96(4), 821–834.

Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V. and Vannucci, M. (2018) A Bayesian approach for estimating dynamic functional network connectivity in fMRI data. *Journal of the American Statistical Association*, 113(521), 134–151.

Yuan, M. and Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1), 19–35.

Zhou, S. (2014) Gemini: graph estimation with matrix variate normal instances. *The Annals of Statistics*, 42(2), 532–562.

Zhu, Y. and Cribben, I. (2018) Sparse graphical models for functional connectivity networks: best methods and the autocorrelation issue. *Brain Connectivity*, 8(3), 139–165.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 1,2.1-2.3,3,4.1-4.2,and 5.1-5.2, are available with this paper at the Biometrics website on Wiley Online Library. The code for implementing the method is also made available in Supporting Information.

How to cite this article: Kundu S, Risk BB. Scalable Bayesian matrix normal graphical models for brain functional networks. *Biometrics*. 2021;77:439–450. https://doi.org/10.1111/biom.13319