Characterising group-level brain connectivity: a framework using Bayesian exponential random graph models

B.C.L. Lehmann[1,2,3], R.N. Henson[3], L. Geerligs[4], Cam-CAN[5], and S.R. White[1]

[1]MRC Biostatistics Unit, University of Cambridge
[2]Big Data Institute, University of Oxford
[3]Department of Statistics, University of Oxford
[4]MRC Cognition and Brain Sciences Unit, University of Cambridge
[5]Donders Institute for Brain, Cognition and Behaviour, Radboud University

Abstract

The brain can be modelled as a network with nodes and edges derived from a range of imaging modalities: the nodes correspond to spatially distinct regions and the edges to the interactions between them. Whole-brain connectivity studies typically seek to determine how network properties change with a given categorical phenotype such as age-group, disease condition or mental state. To do so reliably, it is necessary to determine the features of the connectivity structure that are common across a group of brain scans. Given the complex interdependencies inherent in network data, this is not a straightforward task. Some studies construct a group-representative network, ignoring individual differences, while other studies analyse networks for each individual independently, ignoring information that is shared across individuals. We propose a Bayesian hierarchical framework based on exponential random graph models to characterise the distribution of a entire population of networks. Using resting-state fMRI data from the Cam-CAN project, a study on healthy ageing, we demonstrate how our method can be used to characterise the brain's functional connectivity structure across a group of individuals.

1 Introduction

Brain connectivity analysis aims to understand how spatially distinct regions of the brain interact with each other. The use of networks to model whole-brain connectivity has become increasingly popular in recent years [4]: by treating distinct regions as nodes and the connections between them as edges, researchers have gained new insights into both the structure and function of the brain. To determine the salient features of the brain's connectivity structure, it is necessary to identify which of those are common across individuals.

Given the complex interdependencies inherent in network data, however, this is not a trivial task. Combining connectivity information across participants remains a key challenge. Most existing methods aim to construct a group-representative network. For example, Achard et al. [1] constructed a group-representative mean functional connectivity network by first taking the mean of the individual functional connectivity matrices and then thresholding the resulting matrix. Song et al. [31] took a similar approach using the median of the individual functional connectivity matrices instead. Sinke et

al. [30] constructed a group-representative structural connectivity network from DTI data by keeping those edges which are present in at least 35% of the individual networks. These edge-based methods are computationally convenient, but they ignore the higher-order topological properties present in each participant's network. Another method attempts to represent a whole group by selecting a 'best' individual network in terms of mutual information [20] or Jaccard index [16]. While this approach preserves the topological properties of a single network, in subsequent analyses it may lead to undue weight being placed on features that are not present in many of the remaining networks.

An alternative approach to constructing group-level distribution of brain networks stems from exponential random graph models (ERGMs). An ERGM is a set of parametric statistical distributions on network data (see [25] for a review). The aim of the model is to characterise the distribution of a network in terms of a set of *summary statistics*. These summary statistics are typically comprised of topological features of the network, such as the number of edges and subgraph counts. The summary statistics enter the likelihood via a weighted sum; the weights are (unknown) model parameters that quantify the relative influence of the corresponding summary statistic on the overall network structure and must be inferred from the data. ERGMs are thus a flexible way in which to describe the global network structure as a function of network summary statistics.

ERGMs have been applied successfully to characterise both functional connectivity [28, 29, 22, 10] and structural connectivity [30]. To date, there have been two proposed approaches for using ERGMs in group studies. The first approach constructs a single group network by keeping edges that are present in a minimum of the individual networks, and then fits an ERGM to the group network [30]. As described above, this edge-based approach may average over informative topological structure present in each of the individual networks. In contrast, the second approach fits an ERGM to each individual network and then takes the mean or median of the fitted parameters to represent the group-level connectivity structure [29, 22]. While this is preferable to the first approach, taking a simple mean or median may obscure important information.

In this article, we introduce a novel hierarchical framework to model a population of individual brain networks. This framework uses a Bayesian formulation of the ERGM to pool information across individuals, thus providing an approach to characterise connectivity at a group-level. We demonstrate our method on functional connectivity networks derived from resting-state fMRI scans of participants from the Cam-CAN project, a study on healthy ageing [26].

2 Methods

2.1 Data

The data were collected as part of Phase II of the CamCAN project (www.cam-can.org; [26]). The MRI data were acquired on a 3T Siemens TIM Trio at the MRC Cognition & Brain Sciences Unit, with a 32 channel head-coil. Structural images were acquired using a 1mm3 isotropic, T1-weighted Magnetization Prepared RApid Gradient Echo (MPRAGE) sequence and a 1mm3 isotropic, T2-weighted Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (SPACE) sequence. The fMRI data for the eyes-closed, resting-state run were acquired using a Gradient-Echo (GE) Echo-Planar Imaging (EPI) sequence, and consisted of 261 volumes (lasting 8 min and 40 s). Each volume contained 32 axial slices (acquired in descending order), with slice thickness of 3.7 mm and interslice gap of 20% (for whole brain coverage including cerebellum; TR 1,970 ms; TE 30ms; voxel-size 3 mm x 3 mm x 4.44 mm). EPI fieldmaps with two TEs (5.19ms and 7.65ms) were also ac-

quired. All the raw data from CamCAN Phase II, together with more acquisition details, are available on: http://www.cam-can.org/index.php?content=dataset.

The data were processed using r7219 of the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm), automated with release r5.4 of the Automatic Analysis (AA) pipeline system [9] (http://github.com/automaticanalysis/automaticanalysis/releases/v5.4.0; for overview of pipelines, see [33]) in r2015 of MATLAB (The MathWorks). To obtain a good starting-point for image normalisation, the T1 image was coregistered to the Montreal Neurological Institute (MNI) template using rigid body transformation, and then the T2 image was coregistered to the T1. Both T1 and T2 images were bias corrected, and then combined in a multimodal segmentation to estimate images of each of six tissue classes, including grey-matter (GM), white-matter (WM) and cerebrospinal fluid (CSF). Diffeomorphic registration (DARTEL) was then applied to the GM and WM segments to create group templates, which were in turn transformed to MNI space using a 12-parameter affine transform.

The fMRI images were unwarped using distortion fields estimated from the fieldmaps, and corrected for motion using rigid-body realignment to the mean fMRI image across runs. The different slice acquisition times were corrected by interpolating to the middle slice. The images were rigid-body coregistered to the T1 image and the spatial transformations from that T1 to MNI space (diffeomorphic and affine) were applied to every fMRI image. Residual effects of abrupt motion were reduced by applying wavelet despiking [23]. The mean time series for all voxels within the thresholded WM and CSF segments were calculated to use as later covariates of no interest.

The fMRI time series were then extracted from 90 cortical and subcortical regions of interest (ROIs) from the AAL atlas [35]. ROIs in the cerebellum were not included. The time series for each voxel in each ROI were adjusted for various confounds by taking the residuals from a general linear model (GLM) that included: 1) the time series in WM and CSF segments, 2) the 6 rigid-body motion parameters from the realignment stage, 3) the first-order difference in those motion parameters across successive TRs and 4) a cosine basis set up to cut-off frequency of 0.008Hz (implementing a form of high-pass filter). Second-order (squared) terms for 1), 2) and 3) were also included in the GLM. The autocorrelation in the GLM error was modelled by a family of 8 exponentials with half-lives from 0.5 to 64 TRs (estimated by pooling across voxels within each ROI), and used to prewhiten the time series. The average voxels of the residual time series was then used as a summary measure for each ROI. This approach was based on the optimised pipeline proposed by Geerligs et al. [12].

To illustrate our method, we analyse the twenty youngest individuals in the Cam-CAN study, aged 18-24.

2.2 Network construction

The preprocessed data consists of $N = 90$ ROI time series for each individual. To apply a standard (binary) ERGM, it is necessary to extract a network for each individual. For a given individual i, we first computed the pairwise Pearson correlation between each of the time series, yielding a $N \times N$ correlation matrix $\boldsymbol{C}^{(i)}$. In contrast with the pipeline proposed by Geerligs et al. [12], we did not apply mean regression to the resulting functional connectivity matrices because we accounted for differences in mean connectivity via the thresholding procedure or through the specification of the ERGM.

To each correlation matrix, we then applied a threshold $r^{(i)}$ to produce an $N \times N$ adjacency matrix $\boldsymbol{A}^{(i)}$ with entries:

$$\boldsymbol{A}^{(i)}_{kl} = \begin{cases} 1 & \text{if } \boldsymbol{C}^{(i)}_{kl} \geq r^{(i)} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The adjacency matrix defines a network $\boldsymbol{y}^{(i)}$ with an edge between nodes k and l if and only if $\boldsymbol{A}_{kl}^{(i)} = 1$. The choice of threshold $r^{(i)}$ can have a significant impact on the resulting connectivity structure [37, 36]. The two most common strategies are "absolute" thresholding and "proportional" thresholding. Absolute thresholding picks a common threshold $r^{(i)} = r$ for each individual. In contrast, proportional thresholding ensures that the same number of edges are present in each network by allowing the threshold to vary by individual.

Both thresholding strategies have their drawbacks. Since the number of edges in a network inherently affects the overall network [37, 14], differences in other network metrics may simply be attributable to variations in the overall connectivity. A proportional threshold may thus be preferred in order to keep the number of edges constant across individuals (in both groups) and facilitate comparison of other metrics of interest. On the other hand, a lower correlation value may be less reliable in indicating a functional connection between brain regions. Therefore, by including lower correlations as edges for individuals with lower overall connectivity, a proportional threshold may induce more randomness in the resulting network [36].

Given these issues, we analysed networks constructed using the two different thresholding procedures (absolute and proportional). For both procedures we considered two distinct threshold values, chosen to yield average node degrees of $K = 3$ and $K = 5$. The value of $K = 3$ corresponds to the efficiency-cost optimization (ECO) criterion [11]. This aims to optimise the trade-off between the efficiency of a network and its density, or wiring cost. The value of $K = 5$ has been used in previous applications of ERGMs to neuroimaging [28, 29].

We also constructed group-representative mean (see e.g. [1]) and median networks [31] by first taking the mean and median, respectively, of the correlation matrices across the 20 participants and then thresholding the resulting matrices to yield networks with average node degrees of $K = 3$ and $K = 5$. While it is computationally convenient to analyse a single group-representative network as opposed to each individual network, this edge-based approach does not necessarily provide a good representation of the group's network characteristics.

2.3 Exponential random graph model specification

The probability mass function of a network $\boldsymbol{Y}$ under an ERGM is given by $\pi(\boldsymbol{y}|\theta)$ where

$$\pi(\boldsymbol{y}|\theta) = \frac{\exp\left\{\theta^T s(\boldsymbol{y})\right\}}{Z(\theta)}. \tag{2}$$

Here, $s(\boldsymbol{y})$ is a vector of p network summary statistics, $\theta \in \Theta \subseteq \mathbb{R}^p$ is a vector of p corresponding model parameters that must be estimated from the data and $Z(\theta) = \sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp\left\{\theta^T s(\boldsymbol{y}')\right\}$ is the normalising constant ensuring the probability mass function sums to one. Given data, that is, an observation $\boldsymbol{y}$ of $\boldsymbol{Y}$, the goal is to infer which values of θ best correspond to the data under this distribution. The summary statistics included in a given ERGM represent those configurations expected to appear more frequently or less frequently than in a random graph. In other words, the choice of summary statistics is a modelling decision; it reflects our belief of how the global network structure may be summarised and is driven by the context of the network. The flexibility of ERGMs derives from the range and number of possible summary statistics that can be used [24].

In what follows, we work in the Bayesian paradigm, treating the model parameters θ as random variables. The Bayesian formulation of ERGMs was first suggested By Koskinen [17] and then expanded upon by Caimo & Friel [6]. Through the machinery of Bayesian hierarchical modelling, this

provides a natural framework for describing the joint distribution of a *group* of networks. To fully specify a Bayesian ERGM, we need only augment the definition in (2) with a prior distribution $\pi(\theta)$ for the model parameters. Given an observation $\boldsymbol{y}$ of the network, we can then perform inference by analysing the posterior distribution $\pi(\theta|\boldsymbol{y})$. Through Bayes' rule, we may write the posterior as

$$\begin{aligned}\pi(\theta|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{y}|\theta)\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y})} \\ &= \frac{\exp\left\{\theta^T s(\boldsymbol{y})\right\}\pi(\theta)}{Z(\theta)\pi(\boldsymbol{y})}\end{aligned} \tag{3}$$

where $\pi(\boldsymbol{y}) = \int_\Theta \pi(\boldsymbol{y}|\theta)\pi(\theta)d\theta$ is the model evidence. Although the posterior distribution is generally not available in a closed-form expression, one may obtain samples from the posterior through a Markov chain Monte Carlo method known as the exchange algorithm [21].

For the absolute thresholding, we modelled each individual network as an exponential random graph with three summary statistics that have previously been applied to single brain functional connectivity networks [28, 27]. The three summary statistics used are 1) number of edges, 2) geometrically weighted edgewise shared partners (GWESP), and 3) geometrically weighted non-edgewise shared partners (GWNSP). For the proportional thresholding, we fixed the number of edges and modelled the network using two summary statistics: GWESP and GWNSP.

The number of edges $E = \sum_{i<j} Y_{ij}$ characterises the sparsity of the network. When using an absolute threshold, it is important to include this term in the model in order to account for differences in the network density across individuals. When using a proportional threshold, however, we opted not to include the term in the model, as the number of edges is fixed across individuals. In this case, we constrained the space of networks to those with the same number of edges as those observed. In other words, the model describes a probability distribution over networks with a given average node degree K. Note that this in contrast to previous applications of ERGMs in neuroimaging, which include the edges term in the model despite using a proportional threshold [28, 29].

The geometrically weighted edgewise shared partner (GWESP) statistic of a network $\boldsymbol{y}$ is a measure of network transitivity (or clustering) and is given by:

$$GWESP(\boldsymbol{y}) = e^{\tau_1} \sum_{i=1}^{V-2} \left\{1 - (1 - e^{-\tau_1})^i\right\} EP_i(\boldsymbol{y}) \tag{4}$$

where $EP_i(\boldsymbol{y})$ is the number of *connected* node pairs having exactly i shared partners and $\tau_1 > 0$ is a decay parameter. The decay parameter τ_1 serves to diminish the effect of the network having more higher-order edgewise shared partners relative to lower-order edgewise shared partners. In other words, the increase in GWESP from adding a single edge is smaller if the edge adds a shared partner to connected nodes that already share many partners.

The geometrically weighted non-edgewise shared partner (GWNSP) statistic is similarly defined as

$$GWNSP(\boldsymbol{y}) = e^{\tau_2} \sum_{i=1}^{V-2} \left\{1 - (1 - e^{-\tau_2})^i\right\} NP_i(\boldsymbol{y}) \tag{5}$$

where $NP_i(\boldsymbol{y})$ is the number of *non-connected* node pairs having exactly i shared partners and $\tau_2 > 0$ is a decay parameter. GWNSP is related to global network efficiency; a higher value of GWNSP indicates that non-connected nodes are more likely to have a shared partner.

While it is possible to treat the decay parameters for both geometrically weighted statistics as

extra model parameters (leading to *curved* ERGMs [15]), this increases the computational burden substantially. The decay parameters were therefore fixed at $\tau_1 = \tau_2 = 0.75$, as these values generally resulted in better fitting models [28].

2.4 Hierarchical framework

The Bayesian exponential random graph model described above provides a flexible family of distributions for a *single* network. Our goal is to extend this to a model for a *population* of networks. Our proposed approach is simple: represent each network as a separate ERGM within a Bayesian multilevel (or hierarchical) model. By pooling information across individual networks, this approach allows us to characterise the distribution of the whole population.

We model each individual network $\boldsymbol{Y}^{(i)}$ as an exponential random graph with model parameter $\theta^{(i)}$. Importantly, each individual ERGM must consist of the same set of summary statistics $s(\cdot)$. The probability mass function of each network can then be written

$$\pi(\boldsymbol{y}^{(i)}|\theta^{(i)}) = \frac{\exp\left\{\theta^{(i)T} s(\boldsymbol{y}^{(i)})\right\}}{Z(\theta^{(i)})}, \quad i = 1, \ldots, n. \tag{6}$$

This specifies the data-generating process for each individual network. To obtain a joint distribution for the set of networks, we assume that, conditional on their respective individual-level parameters, the $\boldsymbol{Y}^{(i)}$ are independent. Thus, the sampling distribution for the set of networks $\boldsymbol{Y}$ is simply the product of the individual probability mass functions:

$$\begin{aligned} \pi(\boldsymbol{y}|\boldsymbol{\theta}) &= \prod_{i=1}^{n} \pi(\boldsymbol{y}^{(i)}|\theta^{(i)}) \\ &= \frac{\exp\left\{\sum_{i=1}^{n} \theta^{(i)T} s(\boldsymbol{y}^{(i)})\right\}}{\prod_{i=1}^{n} Z(\theta^{(i)})}. \end{aligned} \tag{7}$$

To complete the model, we need to specify the prior distribution of the individual-level ERGM parameters $\theta^{(1)}, \ldots, \theta^{(n)}$. To this end, we posit a 2-stage multilevel model: assume the $\theta^{(1)}, \ldots, \theta^{(n)}$ are drawn from a common population-level Normal distribution with parameters $\phi = (\mu, \boldsymbol{\Sigma})$, which are also treated as random variables. We write

$$\theta^{(i)} \sim \pi(\cdot|\phi), \quad i = 1, \ldots, n \tag{8}$$

for the population-level distribution. Assuming that, conditional on ϕ, the $\theta^{(i)}$ are independent, we have

$$\pi(\boldsymbol{\theta}|\phi) = \prod_{i=1}^{n} \pi(\theta^{(i)}|\phi). \tag{9}$$

Finally, write $\pi(\phi)$ for the (hyper)prior distribution of ϕ. The joint distribution of $(\boldsymbol{Y}, \boldsymbol{\theta}, \phi)$ can be written as $\pi(\boldsymbol{y}, \boldsymbol{\theta}, \phi) = \pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\phi)\pi(\phi)$. See Figure 1 for a diagrammatic representation of the full hierarchical framework. Details of the prior specification can be found in the Supplementary Material.

2.5 Model fitting

To generate samples from the joint posterior distribution, we devised a novel MCMC algorithm, the details of which can be found in the Supplementary Material. From these samples, we derived

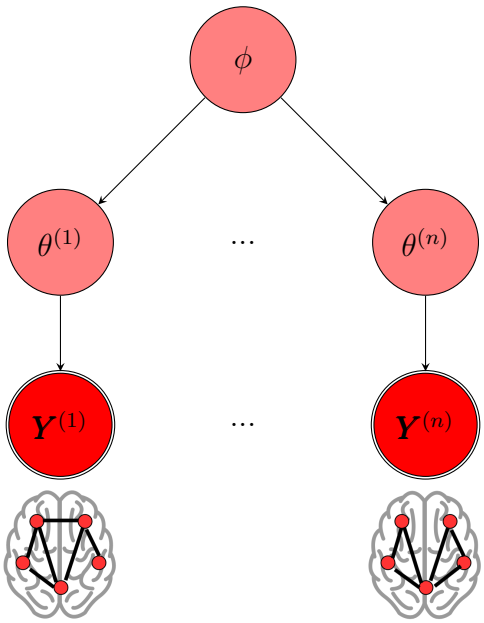


Figure 1: A diagrammatic representation of the hierarchical framework. Each network $\boldsymbol{Y}^{(i)}$ is modelled as an exponential random graph with individual-level parameter $\theta^{(i)}$. In turn, each θ_i is assumed to come from a common population-level distribution with hyperparameter ϕ.

estimates for the posterior mean and 95% credible regions for each pair of components of the group-level parameter μ. In order to compare our method with existing approaches, we also generated posterior samples for single Bayesian ERGM fit to the group-representative mean and median networks, as well as each of the 20 individual networks.

As an alternative, approximative approach to characterising a group-level distribution, we took an average distribution based on the $n = 20$ separate individual ERGM fits. To do so, we first calculated the posterior sample mean $\bar{\theta}^{(i)}$ and posterior sample covariance $\hat{\boldsymbol{\Sigma}}^{(i)}$ for each individual i. We then specified the alternative group distribution to be a Normal distribution $\mathcal{N}(\theta_G, \boldsymbol{\Sigma}_G)$ with mean

$$\theta_G = \frac{1}{n} \sum_{i=1}^{n} \bar{\theta}^{(i)}$$

and covariance

$$\boldsymbol{\Sigma}_G = \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\Sigma}}^{(i)}.$$

2.6 Goodness-of-fit assessment

We assessed the goodness-of-fit of the posterior distribution to the data by simulating networks from the posterior predictive distribution and comparing the network metrics of these simulated networks to those of the observed networks. Specifically, we chose uniformly at random $S = 200$ samples $\mu^{(1)}, \ldots, \mu^{S}$ generated from the posterior distribution of the group-level parameter μ. For each sample, we simulated a network $\boldsymbol{Y}^{(s)}$ from the ERGM with parameter $\mu^{(s)}$. We compared the simulated networks to the observed the networks on three network metrics: degree distribution, geodesic distance distribution (length of shortest paths) and edge-wise shared partners distribution.

We also assessed the goodness-of-fit of the distribution derived directly from the individual network fits. In this case, we generated $S = 200$ samples $\mu_G^{(1)}, \ldots, \mu_G^{S}$ from $\mathcal{N}(\theta_G, \boldsymbol{\Sigma}_G)$, and simulated networks from these parameter values for comparison with the observed data.

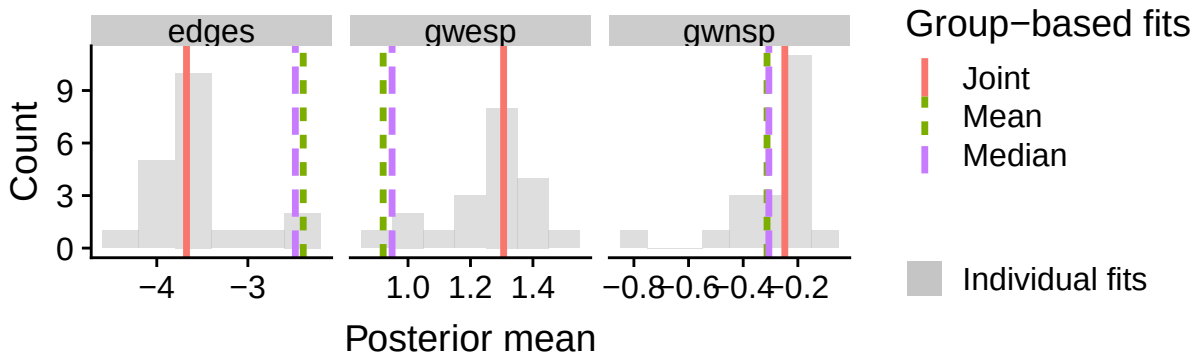


Figure 2: Posterior mean estimates of models fit to networks constructed under absolute thresholding with group-wide average node degree $K = 5$. The histogram (grey bars) corresponds to the posterior mean values of the model parameters fitted to each of the $n = 20$ individual networks. The vertical lines correspond to the posterior mean values of group-level mean parameter under the joint model (red line) and the model parameters fitted to the group-representative mean network (green dotted line) and the group-representative median network (purple dashed line). For each of the model components, the posterior mean under the joint model lay in the centre of the individual network posterior mean values. In contrast, for the edges and gwesp components, posterior means for the mean and median networks lay in the tails of the individual network mean values.

3 Results

We performed analyses for networks constructed under two different thresholding procedures (absolute v. proportional) and two different threshold values (chosen to yield average node degrees of $K = 3$ and $K = 5$), resulting in four sets of networks. For each set of networks, we generated posterior samples from single Bayesian ERGMs to each of the $n = 20$ individual networks as well as the group-representative mean and median networks. We also generated posterior samples from our proposed joint model, which combines information from each individual network simultaneously. Here, we present the results of the analyses performed on the set of networks constructed under absolute thresholding with a group-wide average node degree of $K = 5$. The results under the remaining thresholding procedures can be found in the Supplementary Material.

Comparison to group-representative mean and median networks

We first compare our joint model approach to using group-representative mean and median networks. To do so, we calculated the posterior mean of the group-level parameters estimated from each method and contrasted these with the posterior mean of the parameters corresponding to each of the individual networks (Figure 2). For each of the three components, the posterior mean of the group-level parameter under the joint model (red line) lay roughly in the centre of the posterior mean values across the individual network fits (grey bars). In contrast, the posterior means of the edges and gwesp components for the mean and median network fits (green and purple lines, respectively) lay in the tails of individual network fit posterior mean values. Similar results were observed under the other thresholding procedures (Figure 5). This suggests that the mean and median networks did not accurately capture the topological structure of the majority of the individual networks.

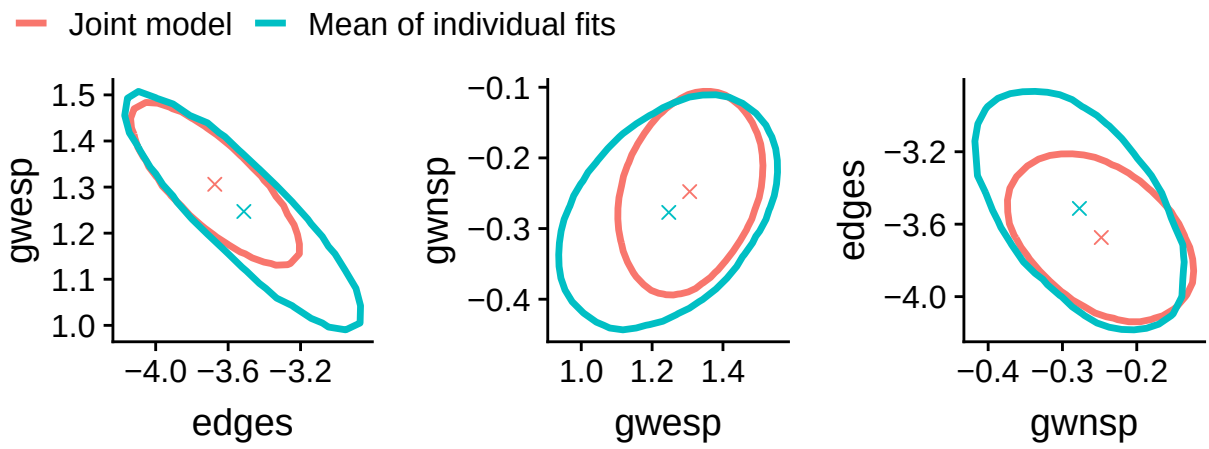


Figure 3: Posterior means and 95% credible regions of the group-level mean parameters for each pair of model components under the joint model (red cross and line) and under the alternative group-level distribution constructed by averaging over the separate individual fits (blue cross and line).

Posterior covariance

Next, we consider the covariance of the posterior distribution under our joint model approach to an approximative model constructed from the mean of the individual fits. From the posterior samples for the group-level mean parameters, we calculated 95% credible regions for each pair of model components. While the posterior mean values were similar across the joint model and the approximative model, the joint model yielded tighter credible regions (Figure 3). The credible regions derived from the other thresholding procedures were similarly tighter under the joint model (Figure 6). This illustrates the one of the benefits of jointly modelling the networks simultaneously: the borrowing of information across individual leads to more precise estimates of the model parameters.

Goodness-of-fit

To assess the goodness-of-fit of the models, we simulated networks from the ERGM with parameters taken uniformly at random from the posterior samples. For both the joint model and the alternative group-level distribution, the simulated networks appeared broadly similar to the observed networks in terms of the network metrics we considered: degree distribution, geodesic distance distribution and edgewise-shared partner distribution (Figure 4). This indicates that the group-level posterior distribution adequately captured the important network characteristics across the entire group of individuals. Similar results were observed under the remaining three thresholding procedures (Figures 7-9).

4 Discussion

The characterisation of connectivity structure at a group level is a necessary step towards understanding how the brain changes with a given phenotype such as age, disease condition or mental state. We have proposed one method of doing so, a Bayesian hierarchical framework based on exponential random graph models. By pooling information across the individual networks, this framework provides a principled approach to model the relational structure for an entire group of networks.

Our work builds on an approach by Simpson et al. to construct a group-representative network from a set of individual networks [29]. In their approach, an ERGM is first fit to each network

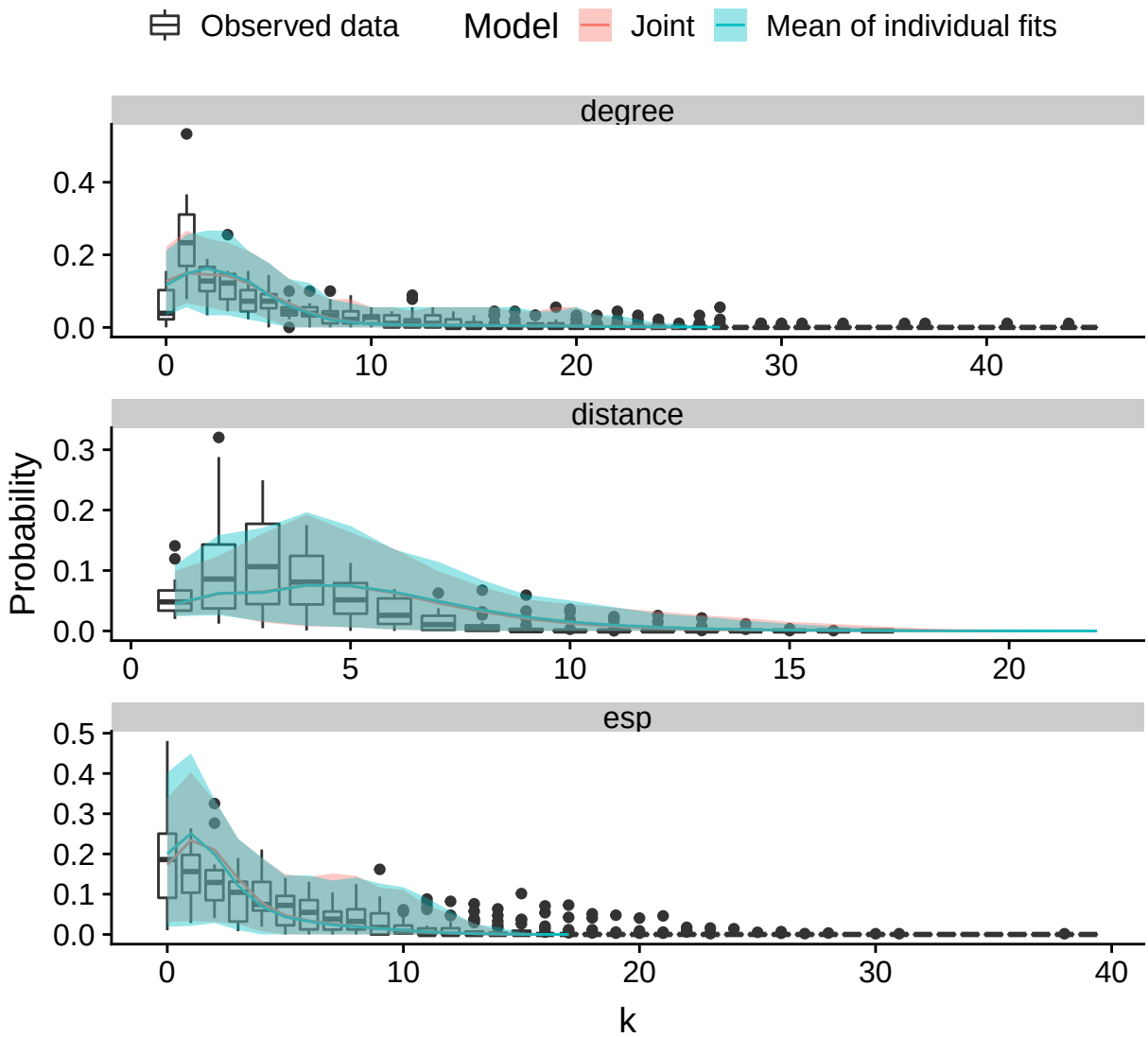


Figure 4: Goodness-of-fit assessment of the joint model (red ribbon) and the alternative group-level distribution constructed by averaging over the separate individual fits (blue ribbon). $S = 200$ networks were generated from the ERGM with model parameters sampled from the respective group-level distributions. The simulated networks are compared to the observed data (shown in the box plots) across three network metrics: degree distribution, geodesic distance distribution, edge-wise shared partner distribution. The coloured lines correspond to the respective means across the simulated networks, while the ribbons correspond to 95% credible intervals.

separately, from which a group-level parameter is calculated as the mean or median of the individual parameter fits across each network. A number of networks are then simulated from the ERGM with the group-level parameter and a single network that best matches the average topological values of the original observed networks is chosen. A similar approach could be used to construct a single group-representative network by simulating networks from the posterior predictive distribution described by our joint model. The advantage of our framework is twofold: first, the Bayesian paradigm allows us to formally quantify uncertainty in the group-level model parameters and second, the hierarchical framework allows information to be borrowed across individuals thus, in principle, improving the accuracy of the parameter estimates.

Our proposed method takes as input any group of networks. While there are a variety of methods to construct networks from functional neuroimaging data, each with their own benefits and disadvantages, we took the most common and simple approach of applying a threshold to the correlation matrices, keeping edges between nodes that exhibited a sufficiently strong positive correlation. The choice of threshold can have a significant impact on the resulting connectivity structure [37, 36] and we found that different threshold values did indeed lead to small differences in the posterior distribution of the model parameter. Regardless of the threshold used, our approach provides a framework to characterise the group-level brain connectivity structure, though further investigation into the effect of thresholding procedure on ERGM parameter estimates is warranted.

From a modelling perspective, the appropriate selection of summary statistics is crucial. This is true of fitting ERGMs to single networks, let alone populations of networks. Moreover, the 'correct' choice likely depends on the network construction method. We used the summary statistics suggested in Simpson et al. [28], which were selected via a graphical goodness of fit approach, in which the 'best' metrics are chosen from a prespecified set of potential metrics. While it is possible to perform Bayesian model selection for an ERGM on a single network [5, 3], further work is needed to extend these approaches to perform model selection for a group of networks.

The computational cost of our fitting procedure is considerable. Using a 20-core computing cluster (Intel Xeon E7-8860 v3, 2.2 GHz), the algorithm took between 8 and 12 hours to produce the joint posterior samples for each of the datasets and models considered in this chapter. (This does not include the time taken to tune the proposal functions.) While the computational cost should increase roughly linearly in the number of networks, Krivitsky & Handcock [19] provide empirical evidence suggesting the cost may grow on the order of $p(|V| + |E|)\log(|E|)$ where p is the number of summary statistics, $|V|$ is the number of nodes, and $|E|$ is the number of edges. This limits the size of the networks that our framework can currently handle to relatively coarse parcellations of the brain (c. 100 regions). Improving the efficiency of fitting ERGMs is an ongoing area of research [2, 32, 8] and our framework will benefit from ongoing developments and may be applicable to more spatially resolved networks in the future.

The flexibility of Bayesian hierarchical modelling could be exploited to deal with more complex group structures such as multiple groups or factorial designs. This would enable the statistical comparison of connectivity structure across groups via the resulting posterior samples produced by the model. Similarly, given multiple networks per individual, it would be straightforward to extend the framework by adding another layer to the model. It would then be possible to apply our framework to dynamic functional connectivity or task-based scans in order to query how connectivity structure changes with time or across mental state. One could also incorporate individual-level covariate information into the model to study how connectivity varies with a continuous phenotype such as age.

References

[1] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and Ed Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.

[2] Lampros Bouranis, Nial Friel, and Florian Maire. Efficient bayesian inference for exponential random graph models by correcting the pseudo-posterior distribution. *Social Networks*, 50(Supplement C):98 – 108, 2017.

[3] Lampros Bouranis, Nial Friel, and Florian Maire. Bayesian model selection for exponential random graph models via adjusted pseudolikelihoods. *Journal of Computational and Graphical Statistics*, 27(3):516–528, 2018.

[4] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, Mar 2009.

[5] A. Caimo and N. Friel. Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11 – 24, 2013.

[6] Alberto Caimo and Nial Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41 – 55, 2011.

[7] Alberto Caimo and Nial Friel. Bergm: Bayesian exponential random graphs in r. *Journal of Statistical Software, Articles*, 61(2):1–25, 2014.

[8] Andressa Cerqueira, Aurélien Garivier, and Florencia Leonardi. A note on perfect simulation for exponential random graph models. *arXiv e-prints*, page arXiv:1710.00873, Oct 2017.

[9] Rhodri Cusack, Alejandro Vicente-Grabovetsky, Daniel J. Mitchell, Conor J. Wild, Tibor Auer, Annika C. Linke, and Jonathan E. Peelle. Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using matlab and xml. *Frontiers in Neuroinformatics*, 8:90, 2015.

[10] John Dell'Italia, Micah A. Johnson, Paul M. Vespa, and Martin M. Monti. Network analysis in disorders of consciousness: Four problems and one proposed solution (exponential random graph models). *Frontiers in Neurology*, 9:439, 2018.

[11] Fabrizio De Vico Fallani, Vito Latora, and Mario Chavez. A Topological Criterion for Filtering Information in Complex Brain Networks. *PLOS Computational Biology*, 13(1):e1005305, January 2017.

[12] Linda Geerligs, Kamen A. Tsvetanov, Cam-CAN, and Richard N. Henson. Challenges in measuring individual differences in functional connectivity using fmri: The case of healthy aging. *Human Brain Mapping*, 38(8):4125–4156, 2017.

[13] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.

[14] Cedric E. Ginestet, Arnaud P. Fournel, and Andrew Simmons. Statistical network analysis for functional mri: summary networks and group comparisons. *Frontiers in Computational Neuroscience*, 8:51, 2014.

[15] David R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29(2):216 – 230, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[16] Karen E. Joyce, Paul J. Laurienti, Jonathan H. Burdette, and Satoru Hayasaka. A new measure of centrality for brain networks. *PLOS ONE*, 5(8):1–13, 08 2010.

[17] JH Koskinen. Bayesian analysis of exponential random graphs-estimation of parameters and model selection. Technical report, Stockholm University, 2004.

[18] Johan H. Koskinen, Garry L. Robins, Peng Wang, and Philippa E. Pattison. Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, 35(4):514–527, October 2013.

[19] Pavel N. Krivitsky and Mark S. Handcock. Supplementary material: A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46, 2014.

[20] David Meunier, Renaud Lambiotte, Alex Fornito, Karen Ersche, and Edward Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3:37, 2009.

[21] Iain Murray, Zoubin Ghahramani, and David J. C. MacKay. MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 359–366. AUAI Press, 2006.

[22] Catalina Obando and Fabrizio De Vico Fallani. A statistical model for brain networks inferred from large-scale electrophysiological signals. *Journal of The Royal Society Interface*, 14(128):20160940, March 2017.

[23] Ameera X. Patel, Prantik Kundu, Mikail Rubinov, P. Simon Jones, Petra E. Vértes, Karen D. Ersche, John Suckling, and Edward T. Bullmore. A wavelet method for modeling and despiking motion artifacts from resting-state fmri time series. *NeuroImage*, 95:287 – 304, 2014.

[24] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173 – 191, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[25] Garry Robins, Tom Snijders, Peng Wang, Mark Handcock, and Philippa Pattison. Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29(2):192 – 215, 2007. Special Section: Advances in Exponential Random Graph (p*) Models.

[26] Meredith A. Shafto, Lorraine K. Tyler, Marie Dixon, Jason R. Taylor, James B. Rowe, Rhodri Cusack, Andrew J. Calder, William D. Marslen-Wilson, John Duncan, Tim Dalgleish, Richard N. Henson, Carol Brayne, and Fiona E. Matthews. The cambridge centre for ageing and neuroscience (cam-can) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, 14(1):1–25, 2014.

[27] Sean Simpson, Robert Lyday, Satoru Hayasaka, Anthony Marsh, and Paul Laurienti. A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience*, 7:171, 2013.

[28] Sean L Simpson, Satoru Hayasaka, and Paul J Laurienti. Exponential random graph modeling for complex brain networks. *PloS one*, 6(5):e20039, 2011.

[29] Sean L. Simpson, Malaak N. Moussa, and Paul J. Laurienti. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *NeuroImage*, 60(2):1117 – 1126, 2012.

[30] Michel R.T. Sinke, Rick M. Dijkhuizen, Alberto Caimo, Cornelis J. Stam, and Willem M. Otte. Bayesian exponential random graph modeling of whole-brain structural networks across lifespan. *NeuroImage*, 135(Supplement C):79 – 91, 2016.

[31] M. Song, Y. Liu, Y. Zhou, K. Wang, C. Yu, and T. Jiang. Default network and intelligence difference. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2212–2215, Sep. 2009.

[32] Linda S. L. Tan and Nial Friel. Bayesian variational inference for exponential random graph models. *arXiv e-prints*, page arXiv:1811.04249, Nov 2018.

[33] Jason R. Taylor, Nitin Williams, Rhodri Cusack, Tibor Auer, Meredith A. Shafto, Marie Dixon, Lorraine K. Tyler, Cam-CAN, and Richard N. Henson. The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144:262 – 269, 2017. Data Sharing Part II.

[34] S. Thiemichen, N. Friel, A. Caimo, and G. Kauermann. Bayesian exponential random graph models with nodal random effects. *Social Networks*, 46:11–28, July 2016.

[35] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, 15(1):273 – 289, 2002.

[36] Martijn P. van den Heuvel, Siemon C. de Lange, Andrew Zalesky, Caio Seguin, B.T. Thomas Yeo, and Ruben Schmidt. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage*, 152:437 – 449, 2017.

[37] Bernadette C. M. van Wijk, Cornelis J. Stam, and Andreas Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PLOS ONE*, 5(10):1–13, 10 2010.

[38] Yaming Yu and Xiao-Li Meng. To center or not to center: That is not the question—an ancillarity–sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

5 Supplementary Material

5.1 Prior specification

The choice of prior distributions for Bayesian ERGMs has yet to be studied in any great detail. The appropriate setting of priors is a challenging task due to the typically high levels of dependence between parameters [18]. Studies thus far have generally assumed (flat) multivariate normal prior distribution on the model parameters [6, 30, 34]. For simplicity, we also assume multivariate normal priors, though alternative specifications fully warrant further investigation. Specifically, we assume

$$\theta^{(i)} \sim \mathcal{N}(\mu, \mathbf{\Sigma}_\theta), \quad i = 1, \ldots, n. \tag{10}$$

Thus, in the above notation, we have $\phi = (\mu, \mathbf{\Sigma}_\theta)$.

For the hyperprior distribution of $\phi = (\mu, \mathbf{\Sigma}_\theta)$, we assume a partially conjugate prior. This assumes μ and $\mathbf{\Sigma}_\theta$ are a priori independent, placing a multivariate normal prior on μ and an inverse-Wishart prior on $\mathbf{\Psi}$:

$$\begin{aligned} \mu &\sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_\mu) \\ \mathbf{\Sigma}_\theta &\sim \mathcal{W}^{-1}(\mathbf{\Psi}_\theta, \nu_\theta) \end{aligned} \tag{11}$$

where $\mu_0 \in \mathbb{R}^p$, $\nu_\theta \in \mathbb{R}$ is such that $\nu_\theta > p-1$, and $\mathbf{\Sigma}_\mu, \mathbf{\Psi}_\theta$ are constant, $p \times p$ positive-definite matrices. These choices of hyperpriors are by simplicity as well as computational convenience.

We set $\nu_\theta = p + 1$ and use an empirical approach to set the hyperparameters $(\mu_0, \mathbf{\Sigma}_\mu, \mathbf{\Psi}_\theta)$. We do this by first fitting a single Bayesian ERGM to each network, using the `Bergm` package [7], with the same summary statistics as in the full multilevel model. For each network $\boldsymbol{y}^{(i)}$, we produced $K = 12000$ samples $\{\theta_k^{(i)}\}_{k=1}^K$ from their respective posterior distributions $\pi(\theta^{(i)}|\boldsymbol{y}^{(i)})$. In particular, we obtained estimates $\bar{\theta}^{(i)} = \frac{1}{K}\sum_{k=1}^K \theta_k^{(i)}$ of the mean and $\hat{\mathbf{\Sigma}}^{(i)} = \frac{1}{K-1}\sum_{k=1}^K (\theta_k^{(i)} - \bar{\theta}^{(i)})(\theta_k^{(i)} - \bar{\theta}^{(i)})^T$ of the covariance for the posterior distribution from each network. We then set $\mu_0 = \bar{\mu}$, $\mathbf{\Sigma}_\mu = \sum_i \hat{\mathbf{\Sigma}}^{(i)}$ and $\mathbf{\Sigma}_\theta = \frac{1}{n-1}\sum_{i=1}^n (\bar{\theta}^{(i)} - \bar{\mu})(\bar{\theta}^{(i)} - \bar{\mu})^T$.

5.2 MCMC details

In order to perform inference through this model, we would like to generate samples from its posterior distribution $\pi(\boldsymbol{\theta}, \phi|\boldsymbol{y})$. This posterior is doubly-intractable so standard MCMC schemes such as the Metropolis algorithm are not suitable. We have developed a novel MCMC scheme, the *exchange-within-Gibbs algorithm* to sample from a doubly-intractable posterior within a hierarchical model. Our algorithm combines ideas from the exchange algorithm [21], Gibbs sampling [13], and the ancillarity-sufficiency interweaving strategy [38] to produce samples from the desired posterior.

We treat the unknown parameters of the model $(\boldsymbol{\theta}, \mu, \mathbf{\Sigma}_\theta)$ as *components* of a single multi-dimensional parameter. The idea behind Gibbs sampling is to iteratively sample each component from its conditional distribution given the remaining components. Specifically, at the kth iteration, first sample

$$\mathbf{\Sigma}_{\theta,k} \sim \pi(\cdot|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1}). \tag{12}$$

Then, draw intermediate updates of the individual-level parameters by performing an exchange step. For each i, draw a proposal $\theta' \sim \mathcal{N}(0, \tau_\theta \frac{1}{n}\sum_{i=1}^n \hat{\mathbf{\Sigma}}^{(i)}$ (where $\tau_\theta > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4) followed by an exponential random graph $\boldsymbol{y}'$ from $\pi(\cdot|\theta')$

and then set $\theta_{k-0.5}^{(i)} = \theta'$ with probability

$$\min\left(1, \frac{\exp\left\{\theta'^T[s(\boldsymbol{y}^{(i)}) - s(\boldsymbol{y}')] - \frac{1}{2}(\theta' - \mu_k)^T \boldsymbol{\Sigma}_{\theta,k}^{-1}(\theta' - \mu_k)\right\}}{\exp\left\{\theta_{k-1}^{(i)T}[s(\boldsymbol{y}^{(i)}) - s(\boldsymbol{y}')] - \frac{1}{2}(\theta_{k-1}^{(i)} - \mu_k)^T \boldsymbol{\Sigma}_{\theta,k}^{-1}(\theta_{k-1}^{(i)} - \mu_k)\right\}}\right) \tag{13}$$

Else, set $\theta_{k-0.5}^{(i)} = \theta_{k-1}^{(i)}$. Next, perform an intermediate update of the group-level parameter μ based on these values:

$$\mu_{k-0.5} \sim \mathcal{N}\left(\left(\boldsymbol{\Sigma}_{\mu}^{-1} + n\boldsymbol{\Sigma}_{\theta,k}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_{\mu}^{-1}\alpha + n\boldsymbol{\Sigma}_{\theta,k}^{(-1)}\bar{\theta}_{k-0.5}\right), \left(\boldsymbol{\Sigma}_{\mu}^{-1} + n\boldsymbol{\Sigma}_{\theta,k}^{-1}\right)^{-1}\right). \tag{14}$$

Following this, we set $\tilde{\theta}_k^{(i)} = \theta_{k-0.5}^{(i)} - \mu_{k-0.5}$ for $i = 1, \ldots, n$. Finally, we perform another update of μ. Namely, we draw a proposal $\mu' \sim \mathcal{N}(\mu, \tau_\mu \frac{1}{n-1}\sum_{i=1}^n(\bar{\theta}^{(i)} - \bar{\mu})(\bar{\theta}^{(i)} - \bar{\mu})^T)$ (where $\tau_\mu > 0$ is a tuning parameter chosen to yield acceptance rates between 0.1 and 0.4) followed by a population of exponential random graphs $\boldsymbol{y}^{(i')}$ from $\pi(\cdot|\mu' + \tilde{\theta}_k^{(i)})$ for $i = 1, \ldots, n$. With probability

$$\min\left(1, \frac{\exp\left\{\mu'^T \sum_{i=1}^n[s(\boldsymbol{y}^{(i)}) - s(\boldsymbol{y}^{(i)\prime})] - \frac{1}{2}\mu'^T\boldsymbol{\Sigma}_{\mu}^{-1}\mu'\right\}}{\exp\left\{\mu_{k-1}^T \sum_{i=1}^n[s(\boldsymbol{y}^{(i)}) - s(\boldsymbol{y}^{(i)\prime})] - \frac{1}{2}\mu_{k-1}^T\boldsymbol{\Sigma}_{\mu}^{-1}\mu_{k-1}\right\}}\right) \tag{15}$$

set $\mu_k = \mu'$. Else, set $\mu_k = \mu_{k-0.5}$. The full procedure is given in Algorithm 1.

Algorithm 1 The exchange-within-Gibbs algorithm for a multilevel Bayesian ERGM

Require: number of MCMC iterations N, initial values $(\mu_0, \tilde{\theta}_0^{(1)}, \ldots, \tilde{\theta}_0^{(n)})$

for $k = 1, \ldots, N$ **do**

- draw $\boldsymbol{\Sigma}_{\theta,k} \sim \pi(\cdot|\mu_{k-1}, \tilde{\boldsymbol{\theta}}_{k-1})$

for $i = 1, \ldots, n$ **do**

- set $\theta_{k-1}^{(i)} = \mu_{k-1} + \tilde{\theta}_{k-1}^{(i)}$
- draw $\theta' \sim h_\theta(\cdot|\theta_{k-1}^{(i)})$
- draw $\boldsymbol{y}' \sim \pi(\cdot|\theta')$
- set $\theta_{k-0.5}^{(i)} = \theta'$ with probability
  $\min\Big(1, AR(\theta', \theta_{k-1}^{(i)}; \mu_{k-1}, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}^{(i)}, \boldsymbol{y}')\Big)$ ▷ See Eq. (13)
- else, set $\theta_{k-0.5}^{(i)} = \theta_{k-1}^{(i)}$

end for

- draw $\mu_{k-0.5} \sim \pi(\cdot|\boldsymbol{\theta}_{k-0.5}, \boldsymbol{\Sigma}_{\theta,k})$

for $i = 1, \ldots, n$ **do**

- set $\tilde{\theta}_k^{(i)} = \theta_{k-0.5}^{(i)} - \mu_{k-0.5}$

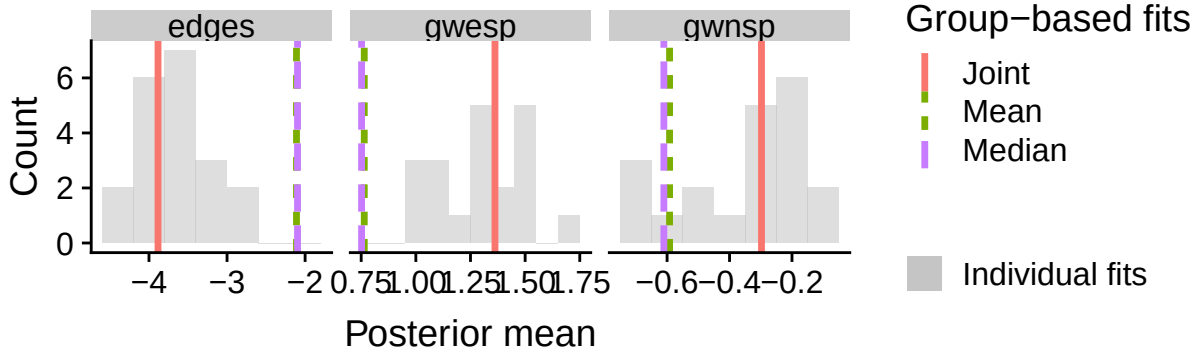
end for

- draw $\mu' \sim h_\mu(\cdot|\mu_{k-0.5})$
- draw $\boldsymbol{y}^{(i)\prime} \sim \pi(\cdot|\mu' + \tilde{\theta}_k^{(i)})$ for $i = 1, \ldots, n$
- set $\mu_k = \mu'$ with probability
  $\min\Big(1, AR(\mu', \mu_{k-0.5}; \tilde{\boldsymbol{\theta}}_k, \boldsymbol{\Sigma}_{\theta,k}, \boldsymbol{y}, \boldsymbol{y}')\Big)$ ▷ See Eq. (15)
- else, set $\mu_k = \mu_{k-0.5}$

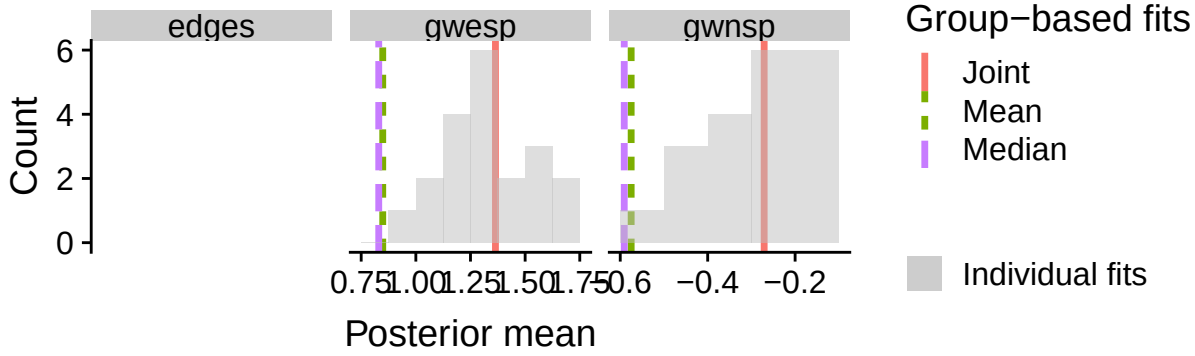
end for

5.3 Supplementary results

(a) Absolute threshold - $k = 3$



(b) Proportional threshold - $k = 3$



(c) Proportional threshold - $k = 5$

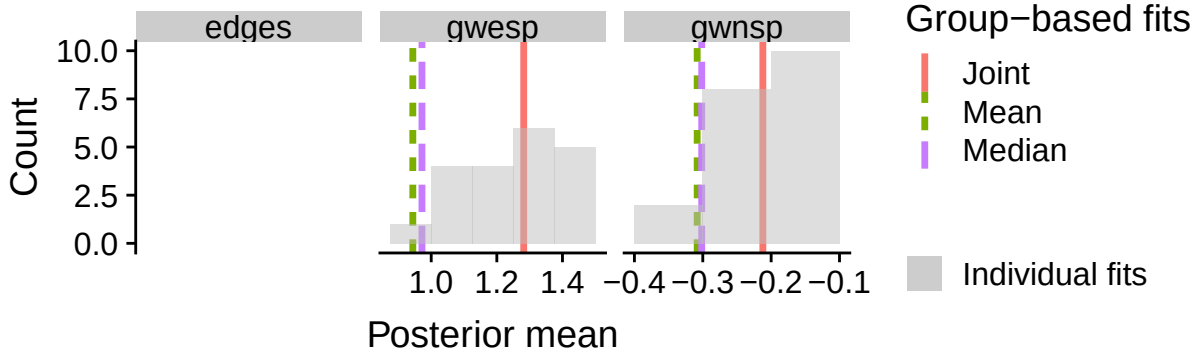
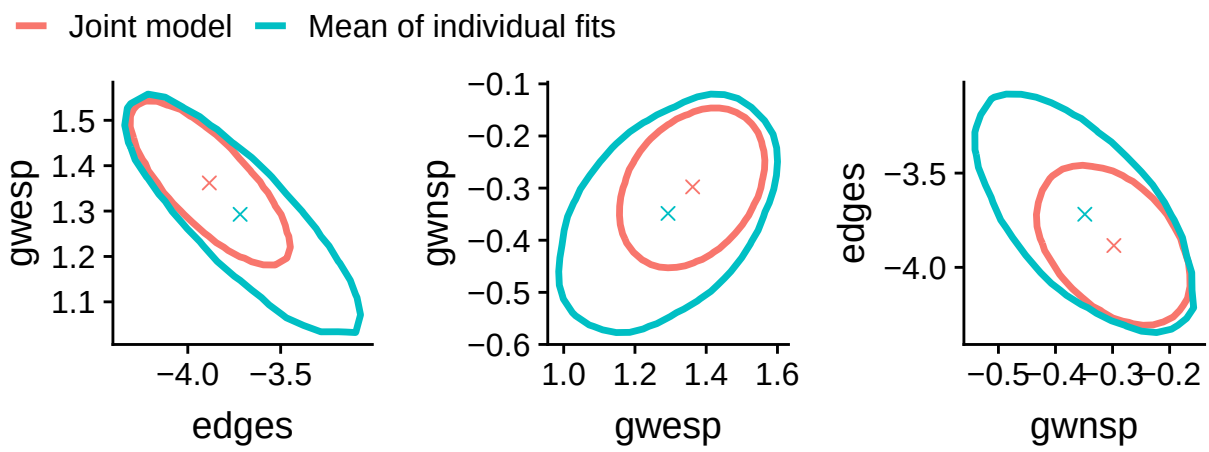
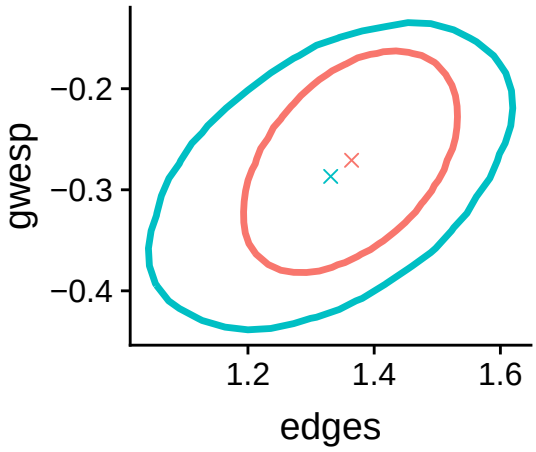


Figure 5: Posterior mean estimates of models fit to networks constructed under a) absolute thresholding with group-wide average node degree $K = 3$, b) proportional thresholding with average node degree $K = 3$ and c) proportional thresholding with average node degree $K = 5$. The histogram (grey bars) corresponds to the posterior mean values of the model parameters fitted to each of the $n = 20$ individual networks. The vertical lines correspond to the posterior mean values of group-level mean parameter under the joint model (red line) and the model parameters fitted to the group-representative mean network (green dotted line) and the group-representative median network (purple dashed line). For each of the thresholding procedures, the posterior mean under the joint model lay in the centre of the individual network posterior mean values. In contrast, posterior means for the mean and median networks lay in the tails of the individual network mean values.

(a) Absolute threshold - $k = 3$



(b) Proportional threshold - $k = 3$



(c) Proportional threshold - $k = 5$

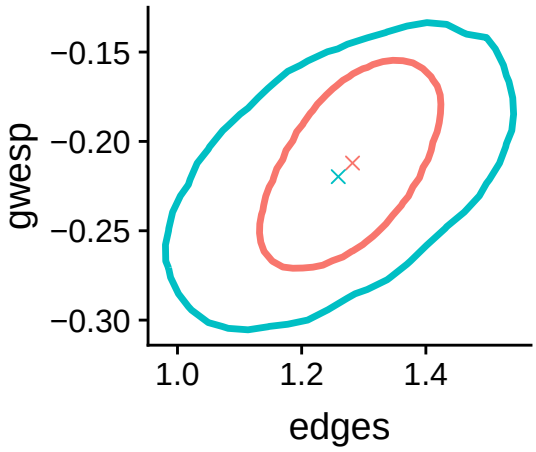


Figure 6: Posterior means and 95% credible regions of the group-level mean parameters for each pair of model components under the joint model (red cross and line) and under the alternative group-level distribution constructed by averaging over the separate individual fits (blue cross and line).

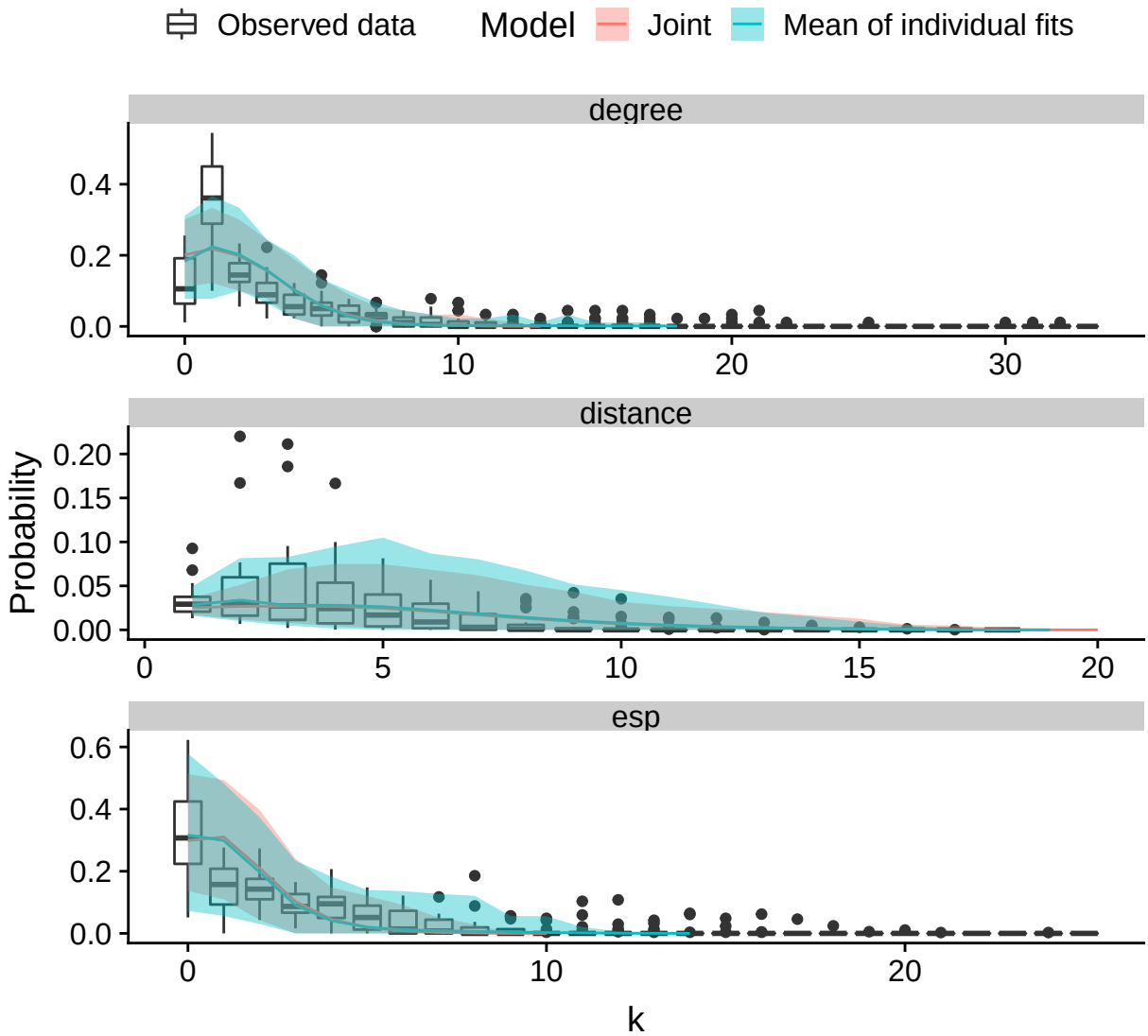


Figure 7: Absolute threshold with group-wide average node degree $K = 3$. Goodness-of-fit assessment of the joint model (red ribbon) and the alternative group-level distribution constructed by averaging over the separate individual fits (blue ribbon). $S = 200$ networks were generated from the ERGM with model parameters sampled from the respective group-level distributions. The simulated networks are compared to the observed data (shown in the box plots) across three network metrics: degree distribution, geodesic distance distribution, edge-wise shared partner distribution. The coloured lines correspond to the respective means across the simulated networks, while the ribbons correspond to 95% credible intervals.

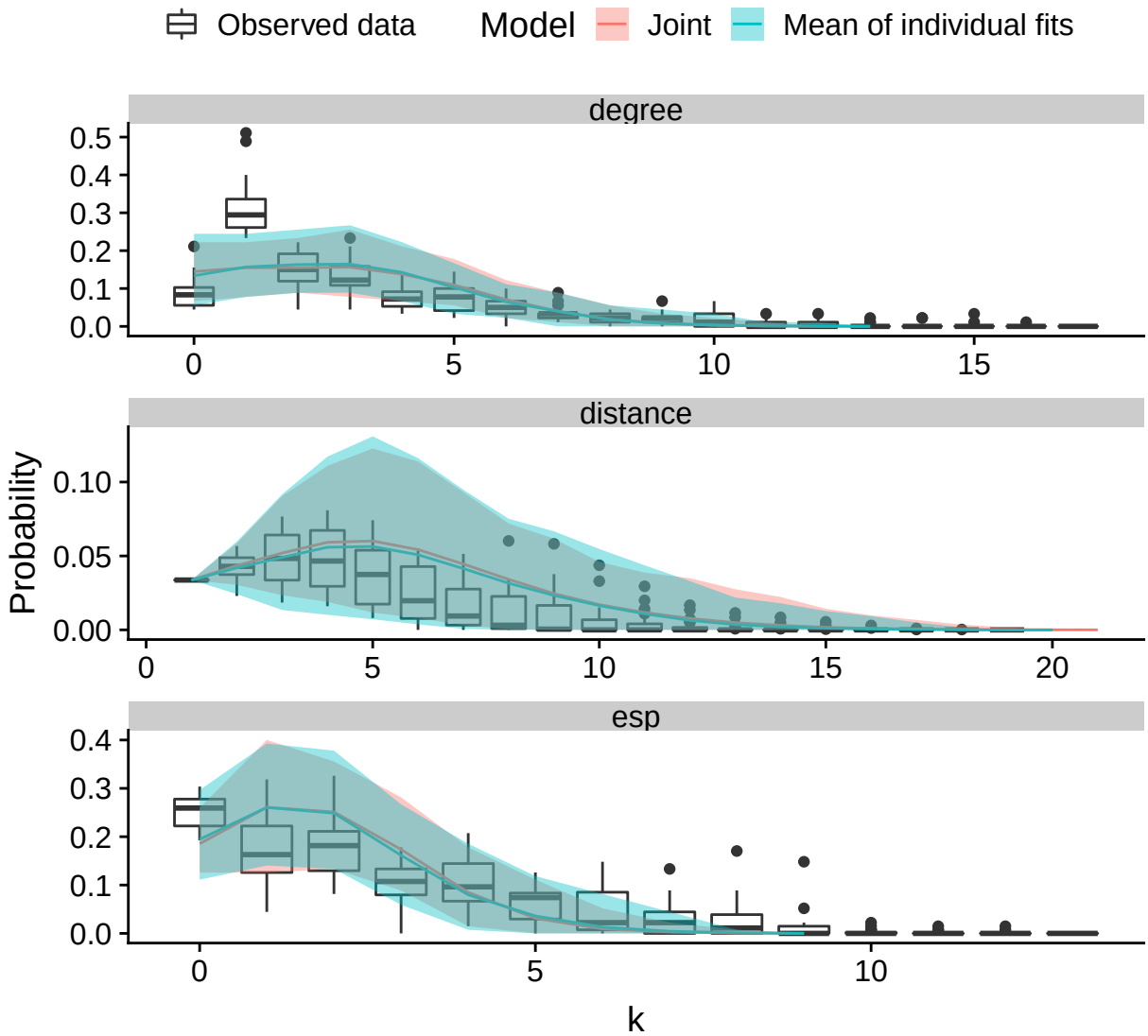


Figure 8: Proportional threshold with average node degree $K = 3$. Goodness-of-fit assessment of the joint model (red ribbon) and the alternative group-level distribution constructed by averaging over the separate individual fits (blue ribbon). $S = 200$ networks were generated from the ERGM with model parameters sampled from the respective group-level distributions. The simulated networks are compared to the observed data (shown in the box plots) across three network metrics: degree distribution, geodesic distance distribution, edge-wise shared partner distribution. The coloured lines correspond to the respective means across the simulated networks, while the ribbons correspond to 95% credible intervals.

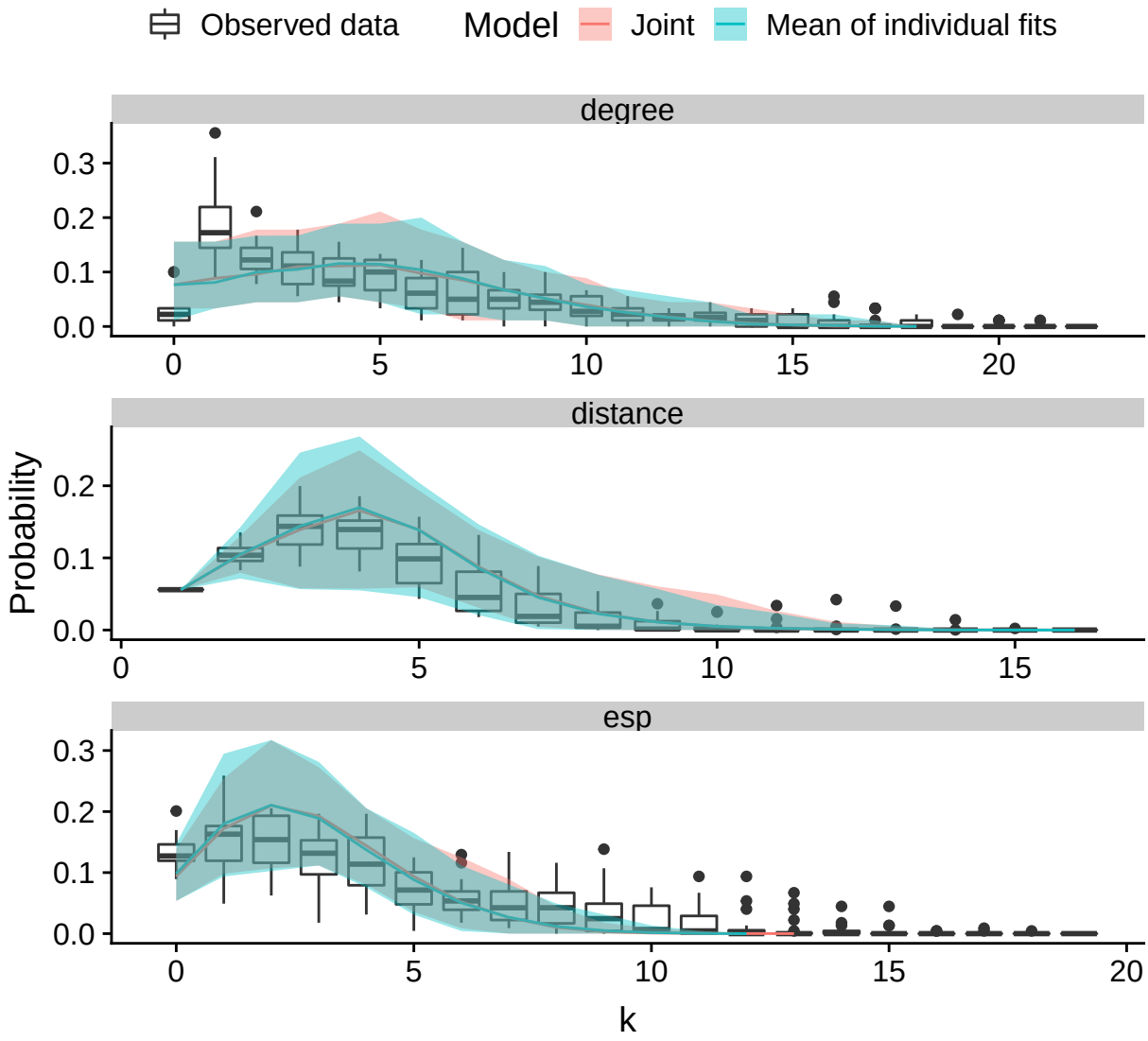


Figure 9: Proportional threshold with average node degree $K = 5$. Goodness-of-fit assessment of the joint model (red ribbon) and the alternative group-level distribution constructed by averaging over the separate individual fits (blue ribbon). $S = 200$ networks were generated from the ERGM with model parameters sampled from the respective group-level distributions. The simulated networks are compared to the observed data (shown in the box plots) across three network metrics: degree distribution, geodesic distance distribution, edge-wise shared partner distribution. The coloured lines correspond to the respective means across the simulated networks, while the ribbons correspond to 95% credible intervals.