advances.in/psychology

May 30, 2024
Doi: 10.56296/aip00016
ISSN: ISSN 2976-937X

METHODS | SPECIAL ISSUE: NETWORK PSYCHOMETRICS

Sensitivity analysis of prior distributions in Bayesian graphical modeling: Guiding informed prior choices for conditional independence testing

Nikola Sekulovski[1*], Sara Keetelaar[1], Jonas Haslbeck[1,2], & Maarten Marsman[1]

Received: September 29, 2023| **Accepted:** May 21, 2024| **Published:** May 30, 2024| **Edited by:** Hudson Golino

[1]Department of Psychology, University of Amsterdam, [2]Department of Clinical Psychological Science, Maastricht University
* Please address correspondence to Nikola Sekulovski, University of Amsterdam, Psychological Methods, Nieuwe Achtergracht 129B, PO Box 15906, 1001 NK Amsterdam, The Netherlands. n.sekulovski@uva.nl.

Bayesian analysis methods provide a significant advancement in network psychometrics, allowing researchers to use the edge inclusion Bayes factor for testing conditional independence between pairs of variables in the network. Using this methodology requires setting prior distributions on the network parameters and on the network's structure. However, the impact of both prior distributions on the inclusion Bayes factor is underexplored. In this paper, we focus on a specific class of Markov Random Field models for ordinal and binary data. We first discuss the different choices of prior distributions for the network parameters and the network structure, and then perform an extensive simulation study to assess the sensitivity of the inclusion Bayes factor to these distributions. We pay particular attention to the effect of the scale of the prior on the inclusion Bayes factor. To improve the accessibility of the results, we also provide an interactive Shiny app. Finally, we present the R package simBgms, which provides researchers with a user-friendly tool to perform their own simulation studies for Bayesian Markov Random Field models. All of this should help researchers make more informed, evidence-based decisions when preparing to analyze empirical data using network psychometric models.

Keywords: Bayesian model averaging; Bayes factor; Bayesian variable selection; Markov random fields; network psychometrics; prior sensitivity

1. INTRODUCTION

Network psychometrics is a popular methodology that uses graphical models to analyze the network structure of multivariate psychological data (Borsboom et al., 2021). In these networks, psychological variables (e.g., symptoms of mental disorders or item responses) are represented by nodes, and the edges between these nodes indicate partial associations between pairs of variables. Markov random field models (MRFs, Kindermann & Snell, 1980; Rozanov, 1982), such as the Ising model for binary variables and the Gaussian Graphical Model for continuous variables, are a specific class of graphical models used in network psychometrics. Their popularity is partly due to their ability to model the underlying conditional independence structure of the data (Lauritzen, 2004). The structure of the MRF reveals how variables are associated once the influence of other variables is taken into account. In other words, the absence of an edge between two variables means that they are conditionally independent after controlling for the remaining variables in the network — the remaining variables explain their direct interaction (Kindermann & Snell, 1980). Several approaches have been proposed in the psychometric literature to analyze MRFs, with the frequentist methodology combining Lasso estimation (Tibshirani, 1996) with EBIC variable selection (Chen & Chen, 2008) being by far the most popular (e.g., Epskamp & Fried, 2018; Friedman et al., 2008; van Borkulo et al., 2014). However, the frequentist approaches have been criticized for their inability to quantify uncertainty about network parameters and structure (Williams, 2021b), which includes the problem of distinguishing between the *absence of evidence* and the *evidence of the absence* of edges within the network (see, for instance, Epskamp et al., 2017; Marsman & Rhemtulla, 2022).

To address the problem of adequately testing for conditional independence, Bayesian methods have been developed to analyse MRF models (Marsman et al., 2015, 2022, 2023a; Mohammadi & Wit, 2015; Williams, 2021a; Williams & Mulder, 2020a). The Bayesian approach provides three methods for testing conditional independence (see Sekulovski et al., 2024 for a detailed overview). One of these methods uses a Bayes factor (Etz & Wagenmakers, 2017; Jeffreys, 1935; Kass & Raftery, 1995) based on Bayesian model averaging techniques (BMA, George, 1999; Hinne et al., 2020; Hoeting et al., 1999) called the inclusion Bayes factor (Huth et al., 2023; Marsman et al., 2023a). This method requires researchers to specify two sets of prior distributions, one for the network structure (i.e., the configuration of present edges) and another for the parameters indicating the strength of the edge weights. As will become clear in this paper, the reason for specifying two sets of priors is because the structure of the network dictates which edge weight parameters are given a particular prior distribution and which are set to zero a priori. Despite the well-developed statistical literature on the foundations of these methods, little attention has been paid to intuitively explaining and illustrating what these prior distributions mean and how they affect the values of the inclusion Bayes factor (but see Huth et al., 2023 for an accessible general introduction to the Bayesian analysis of MRFs). Currently, we lack a clear holistic depiction of how these priors operate under various aspects of the data, such as the sample size,

the number of variables, and the density of the network, among others. The lack of literature that adequately explains and illustrates the aforementioned problems makes it difficult for researchers to understand the influence these prior distributions have on the inferences they make; a problem that discourages them from using this new methodology.

In this paper, we provide statistical and intuitive explanations of the prior distributions for a specific MRF model for ordinal and binary data (Marsman et al., 2023a). We chose to focus on this MRF because ordinal data are arguably the most common data type in cross-sectional applications of network psychometrics; however, we also give a brief overview of the priors used in other popular R packages for analyzing Gaussian graphical models. Since the Bayesian analysis itself requires computational methods, evaluating how prior specifications affect inference in different settings requires a simulation approach. Taking into account different aspects of the data, such as the number of observations, the number of variables, the number of ordinal categories, and the density of the network structure, we conduct an extensive simulation study to demonstrate how prior distribution specifications for both the edge weight parameters and the network structure in the ordinal MRF affect the inclusion Bayes factor when assessing the evidence for conditional (in)dependence. This paper aims to illustrate the interplay between priors and properties of the data at hand and to provide practical guidance so that applied researchers can better exploit the advantages of the Bayesian approach to network psychometrics.

The remainder of this paper is structured as follows. In the next section, we provide a general introduction to the MRF models employed in network psychometrics, with particular attention to the ordinal MRF, and discuss the Bayesian analysis of these graphical models. In the third section, we give the statistical and intuitive background of the prior choices implemented in the analysis of the ordinal MRF, we also briefly review the prior choices that have been implemented or suggested by other authors in this context. In the fourth and fifth sections, we explain the simulation design and highlight some of the main results. In the sixth section, we briefly introduce the new R package simBgms. We conclude the paper with a discussion in which we briefly summarise the main conclusions.

2. GRAPHICAL MODELING

There are two sets of parameters that govern the structure of a MRF model: (i) the configuration of (unweighted) edges, which indicate whether two variables are directly connected; (ii) the corresponding edge weights for the present edges. For example, let's consider the comorbidity between major depressive disorder (MDD) and generalized anxiety disorder (GAD). Assume researchers have collected data on a range of symptoms associated with these disorders, such as low mood, loss of interest, excessive worry, restlessness and sleep problems. By analysing the data using MRF graphical models, the researchers can uncover the conditional independence relationships between these symptoms, revealing which symptoms are directly related to each other, after accounting for the

remaining variables. The resulting network structure (i.e., the configuration of present *unweighted* edges) can provide insight into the common symptoms and potential causal pathways between MDD and GAD (but see Ryan et al., 2022 for a discussion of the limitations of using MRFs for causal discovery). Then, the estimated values of the edge weight parameters provide a measure of the strength and direction (positive/negative) of the relationship between two connected variables, after controlling for the remainder of the variables in the network. This information can guide clinical practice by helping clinicians identify symptoms for treatment and understand the complex interplay between different mental disorders (but also see Dablander & Hinne, 2019). Beyond clinical psychology (see Robinaugh et al., 2020 for a review of the use of network models in psychopathology research), these models can be effectively employed for data-driven theory construction in various domains, such as personality, intelligence, or social psychology research, among others (Borsboom, 2022).

From the two sets of parameters that govern the structure of the MRF model it follows that researchers are implicitly interested in answering two questions: (i) is there an edge between two variables; (ii) what is the strength of the edge weights associated with the existing edges? Although not explicitly stated, the former is the domain of hypothesis testing and the latter is the domain of parameter estimation. By viewing the analysis of graphical models in this way, it becomes obvious that we need to account for two sources of uncertainty: (i) uncertainty about the network structure $\mathcal{S}$ (i.e., the configuration of (unweighted) present edges); and (ii) uncertainty about the parameters θ_{ij} representing the estimated edge weight for a present edge between variables i and j. In our view, applied researchers and some methodologists tend to ignore the first source of uncertainty and focus only on the second. This is usually done by assuming a single network structure a priori, even if this is not explicitly stated, a discussion we will return to later in the paper.

2.1 The Ordinal MRF

Depending on the level of measurement of the variables included in the network, different MRFs could be used to model the dependency structure of the variables. Popular examples are the Gaussian graphical model for continuous data (Epskamp et al., 2018; Lauritzen, 2004; Mohammadi & Wit, 2015) and the Ising model for binary data (Cox, 1972; Ising, 1925; Marsman et al., 2022; van Borkulo et al., 2014). Until recently, despite the abundance of ordinal data in psychological research (largely due to the ubiquitous use of Likert scales in questionnaires) researchers using network psychometric methods were usually forced to treat ordinal data as continuous and fit Gaussian graphical models. However, misspecifying the level of measurement of the data, although common practice in psychology, has been shown to lead to unreliable parameter and uncertainty estimates (Johal & Rhemtulla, 2021; Liddell & Kruschke, 2018). Marsman et al. (2023a) have recently established a MRF for ordinal data and proposed a Bayesian model averaging approach for its estimation, allowing researchers to (i) estimate the models and appropriately test for the presence of edges and (ii) do so at the appro-

priate level of measurement. In the remainder of this subsection, we briefly introduce the ordinal MRF.

Let **X** be a vector of ordinal variables X_i for $i = 1, \ldots, p$ with possible realizations $x_i = 0, \ldots, m_i$, i.e., each variable has $m_i + 1$ ordinal response categories.[1] The ordinal MRF (Marsman et al., 2023a) for **X** is then defined as the joint distribution of the data

$$p(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Theta}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{p} \sum_{h=1}^{m_i} \mathcal{I}(X_i = h)\, \mu_{ih} + 2 \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} X_i X_j \theta_{ij} \right), \tag{1}$$

where μ_{ih} denotes the category threshold for category h of the variable i, with $i = 1, ..., p$ and $h \in \{1, ..., m_i\}$; with $\mathcal{I}(.)$ denoting an indicator function which is 1 if $x_i = h$ and zero otherwise. The category threshold parameters in $\boldsymbol{\mu}$ are often treated as a nuisance in the network psychometrics literature; however, these parameters capture the remaining tendency of the ordinal variables towards a certain category that cannot be explained away by conditioning on the remaining variables in the network.[2] The edge weight parameters θ_{ij}, stored in the symmetric matrix $\boldsymbol{\Theta}$, are of primary interest to researchers using MRF models. Note also that when $m = 1$ the ordinal MRF in Equation 1 is equivalent to the Ising model (Cox, 1972; Ising, 1925). Finally,

$$Z = \sum_{\mathbf{x} \in \mathcal{X}} \exp\left(\sum_{i=1}^{p} \sum_{h=1}^{m_i} \mathcal{I}(X_i = h)\, \mu_{ih} + 2 \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} X_i X_j \theta_{ij} \right),$$

is the normalizing constant, with $\mathcal{X}$ denoting the space of all possible response patterns. For example, when all the variables in **X** have m ordinal categories, the normalizing constant Z has m^p terms, making it computationally infeasible in most applied situations. Therefore, the authors resort to using the pseudolikelihood (further denoted as p^*, Besag, 1974) as an approximation to the full model. Keetelaar et al. (2024) show that, for the Ising model, the maximum pseudolikelihood estimates are comparable to those based on the full likelihood.

Marsman et al. (2023a) show that the MRF in Equation 1 is equivalent to two other graphical models for ordinal data proposed in the psychometrics literature by Anderson and Vermunt (2000) and in the machine learning literature by Suggala et al. (2017); assuring that this MRF preserves the ordinal nature of the variables.

2.2 Bayesian Analysis of MRF Models in Network Psychometrics

In a network of p variables, there can be a maximum of $k = p(p-1)/2$ edges, resulting in 2^k structures as potential statistical models. This leads to a large number of models even for a small number of variables, which means that we quickly become quite uncertain about the true structure. It is there-

[1] Observe that each response variable is allowed to have a different number of ordinal categories.
[2] The threshold parameters play an important role in computing the dynamic landscape of the network.

fore essential to consider all these possible combinations of present and absent edges to adequately account for the structure uncertainty.

The Bayesian Model Averaging (BMA) framework is a powerful approach to account for uncertainty in both the network structure and the associated parameters (e.g., Hinne et al., 2020; Hoeting et al., 1999), achieved by introducing the two sets of prior distributions. Then, utilizing the available data, we update and refine the initial uncertainty represented by the prior distributions, resulting in a joint posterior distribution of the form

$$\underbrace{p(\boldsymbol{\mu}, \boldsymbol{\Theta}, \mathcal{S} \mid \mathbf{X})}_{\text{joint posterior}} \propto \underbrace{p^*(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Theta}, \mathcal{S})}_{\text{pseudolikelihood}} \times \underbrace{p(\boldsymbol{\mu})}_{\substack{\text{prior on the} \\ \text{category thresholds}}} \times \underbrace{p(\boldsymbol{\Theta} \mid \mathcal{S})}_{\substack{\text{prior on the} \\ \text{edge weights}}} \times \underbrace{p(\mathcal{S})}_{\substack{\text{prior on the} \\ \text{graph structure}}} . \tag{2}$$

Equation 2 clearly illustrates why two sets of priors must be specified; $p(\mathcal{S})$, the prior on the network structure (also called the network or graph topology), which in turn determines the prior on the edge weight parameter $p(\boldsymbol{\Theta} \mid \mathcal{S})$. This will become even more obvious in the next section, when we discuss the specific choices for both sets of priors.

Inferring conditional independence relationships is one of the main objectives when estimating the MRF model. Therefore, we wish to test for the presence of an edge between variables i and j, that is, we wish to compare the hypothesis $\mathcal{H}_1^{(ij)}$, which states that there *is* an edge between variables i and j, with the null hypothesis $\mathcal{H}_0^{(ij)}$, which states that there is *no* edge between these two variables. By taking the posterior distribution of the possible structures given the data $p(\gamma \mid \mathbf{X})$, we can obtain posterior inclusion probabilities for every edge. The posterior inclusion probability of the edge between variables i and j can be expressed as the sum over the posterior probabilities of all structures $\mathcal{S}'$ that contain an edge between variables i and j, i.e., for which $\mathcal{H}_1^{(ij)}$ is true

$$p(\mathcal{H}_1^{(ij)} \mid \mathbf{X}) = \sum_{\mathcal{S}' \in \mathcal{S}^{(i-j)}} p(\mathcal{S}' \mid \mathbf{X}).$$

For example, one could include all edges with a posterior inclusion probability greater than 0.5, which would result in the median probability model (Barbieri & Berger, 2004). However, as mentioned above, we wish to *test* $\mathcal{H}_1^{(ij)}$ against $\mathcal{H}_0^{(ij)}$ and one way to do this is to compute the inclusion Bayes factor by contrasting the posterior inclusion odds against the prior inclusion odds (Huth et al., 2023; Marsman et al., 2023a; Sekulovski et al., 2024)

$$\text{BF}_{10} = \underbrace{\frac{p(\mathcal{H}_1^{(ij)} \mid \mathbf{X})}{p(\mathcal{H}_0^{(ij)} \mid \mathbf{X})}}_{\substack{\text{Posterior} \\ \text{inclusion odds}}} \Big/ \underbrace{\frac{p(\mathcal{H}_1^{(ij)})}{p(\mathcal{H}_0^{(ij)})}}_{\substack{\text{Prior} \\ \text{inclusion odds}}} . \tag{3}$$

We can interpret this inclusion Bayes factor as the strength of the evidence in the data in favor of

edge inclusion ($\mathcal{H}_1^{(ij)}$), or take the inverse $1/\text{BF}_{10}$ to quantify the evidence in favor of edge exclusion ($\mathcal{H}_0^{(ij)}$). Although there are no strict cut-off values for the Bayes factor, the current literature on the subject recommends the use of $\text{BF}_{10} > 10$ as strong evidence for edge inclusion (Huth et al., 2023; Jeffreys, 1961). However, researchers are free to decide which value (if any) to choose depending on their substantive research interests (for further guidance see, Kass & Raftery, 1995). If we assume equal prior inclusion odds by giving equal prior probabilities to structures that include the edge and structures that exclude the edge (which we will explain in the following section where we discuss the specific choices for these priors), then the inclusion Bayes factor (BF_{10}) simplifies to the fraction of the posterior edge inclusion probability divided by one minus this probability.

The prior distributions play a crucial role in the outcome of the Bayesian analysis, since they influence the joint posterior distribution (Equation 2), which is used for computing the inclusion Bayes factors (Equation 2). Therefore, it is crucial to thoroughly investigate and understand the prior sensitivity for conditional independence testing when using the ordinal MRF.

3. PRIOR DISTRIBUTIONS

Prior distributions are an essential component of Bayesian inference (e.g., Berger, 1990; Wagenmakers et al., 2018); which means that their selection and specification, whether for parameter estimation or hypothesis testing, requires careful consideration (e.g., Bayarri et al., 2012; Berger, 1990; Consonni et al., 2018). The influence of the prior distributions for the parameters and the prior probabilities of the models in BMA has received considerable attention; but has mostly been limited to variable selection approaches in the familiar domain of regression models (e.g., Eicher et al., 2007; Fernandez et al., 2001; Ley & Steel, 2009). Most extensions of these approaches to other (non-linear) models have been built on the basis of what is known for the linear case; MRFs are no exception.

In the context of BMA analysis of MRFs, specific choices of prior distributions allow us to assign a value of zero to insignificant effects (edge weights) using a variety of variable selection-based techniques, including shrinkage priors or spike and slab priors (Vanucci, 2022). Therefore, the choice of priors should be based on careful evaluation and a solid understanding of the problem at hand. This requires the researcher using the model to have a good understanding of the behavior of the chosen prior distributions. Due to the relatively recent development of BMA approaches for estimating MRFs in the network psychometrics literature, a thorough prior analysis does not yet exist. In this paper, we address this methodological gap.

In the remainder of this section, we provide a comprehensive overview of the prior distributions included in the current implementation of the ordinal MRF model in the R package bgms (Marsman et al., 2023b). In addition to introducing each of these prior options, we provide a brief summary of alternative approaches proposed in the statistical literature and implemented in other R packages.

3.1 Prior Setup for the Ordinal MRF

The fact that there are 2^k possible structures makes the BMA approach to estimating the ordinal MRF model a challenging task for even a small number of variables. Marsman et al., 2023a use a Bayesian variable selection approach (Tadesse & Vanucci, 2022) and hierarchically model the inclusion of individual edges in the network by introducing a latent indicator variable γ and a two-component spike and slab prior (George & McCulloch, 1993; Mitchell & Beauchamp, 1988) for the edge weight parameters. In other words, the method first checks whether an edge exists (structure uncertainty), and if so, assigns it a prior distribution to estimate the value of the edge weight (parameter uncertainty). Therefore, the expression for the joint distribution given in Equation 2 can be rewritten as

$$p(\boldsymbol{\mu}, \boldsymbol{\Theta}, \boldsymbol{\gamma} \mid \mathbf{X}) \propto p^*(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Theta}, \boldsymbol{\gamma}) \times p(\boldsymbol{\mu}) \times p(\boldsymbol{\Theta} \mid \boldsymbol{\gamma}) \times p(\boldsymbol{\gamma}). \tag{4}$$

We can quickly see that the only difference from Equation 2 is that γ is now used to denote the structure $\mathcal{S}$. Note that γ can be viewed as a vector containing a particular configuration (structure) of present and absent edges (i.e., $\boldsymbol{\gamma} = \gamma_{12}, ..., \gamma_k$). Given the setup in Equation 4, we can express the conditional prior of a single edge weight θ_{ij} as follows

$$p(\theta_{ij} \mid \gamma_{ij}) = (1 - \gamma_{ij})\, f_{\text{spike}}(\theta_{ij}) + \gamma_{ij}\ \ f_{\text{slab}}(\theta_{ij}). \tag{5}$$

When the latent indicator variable (γ_{ij}) is zero (i.e., the edge is absent from the network), the edge weight is set to zero by being assigned a spike component; and when the indicator variable is one, the edge weight is assigned a diffuse prior distribution — the slab component. This spike and slab prior strikes a balance between parameter shrinkage (encouraging most parameters to be close to zero) and variable selection (allowing only a subset of parameters to take non-zero values). This is particularly useful for estimating MRF models since it allows us to consider many possible combinations of present and absent edges given the data.

Depending on the choice for f_{spike}, there are two variants of the spike and slab prior. In the first case, f_{spike} is discrete, with its total probability mass at zero (i.e., a Dirac measure that concentrates all of its probability mass at a single value). This simply means that if $\gamma_{ij} = 0$, then the corresponding edge weight θ_{ij} is automatically set to zero. This option is used for the analysis of the ordinal MRF proposed by Marsman et al. (2023a) and implemented in the R package bgms. The second variant is achieved by giving the slab component a highly peaked continuous distribution. This option has been implemented by Marsman et al. (2022) for edge selection in the Ising model for binary data by setting the slab component to a normal distribution with a small variance. A problem with the continuous spike and slab prior is that the intersection between the spike and slab components, set by tuning the variance of the two components is not trivial and can lead to inconsistent structure

selection (see Ly & Wagenmakers, 2021; Marsman et al., 2022).

Since the ordinal MRF uses the discrete spike and slab prior, we only need to consider the distribution for the slab component. Below we discuss three candidate distributions that we implement in the simulation study, one of which is the slab component implemented in the current version of the R package bgms. We also briefly discuss other options for the slab component that we do not implement in the simulation. As can be seen in Equation 5, the spike and slab prior is tied to the value of the latent indicator variable γ_{ij}; therefore, in the third subsection, we elaborate on the two options for these priors and comment on other options available in the literature that we do not implement in the simulation study. We conclude this section with a brief discussion on the priors for the category threshold parameters.

3.2 Prior Distributions for the Edge Weights — The Slab Component

3.2.1 Unit Information

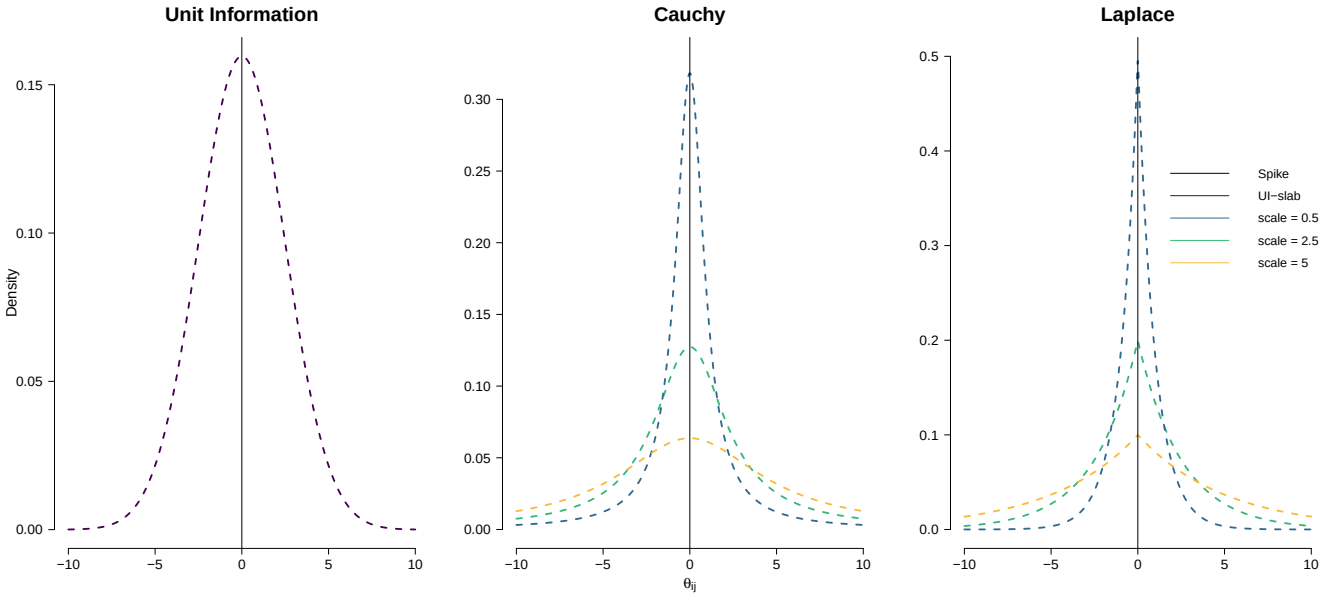
The unit information (UI) prior was originally proposed by Kass and Wasserman (1995) as a Bayes factor approximation to the BIC (Consonni et al., 2018).[3] Its name comes from the fact that this prior approximately contains the information about θ_{ij} in a single observation. It is used when researchers have limited prior information about the parameters in their model. This prior is weakly informative and requires no input from the researcher since it is fully specified by the data. For the analysis of the ordinal MRF, this slab component for each θ_{ij} (left panel of Figure 1) is defined using a normal distribution with a mean of zero and a variance equal to its corresponding value in the diagonal of the covariance matrix of the parameters (i.e., the inverse of the Fisher information matrix, Ly et al., 2017). We also considered an extension of this prior, that takes into account the covariance of the edge weight parameters by including the off-diagonal elements of the covariance matrix. However, due to computational difficulties, we do not include this conditional UI slab component in the simulation study; an explanation of how this prior is defined is given in Appendix A.

3.2.2 Cauchy

The Cauchy distribution has been proposed in the Bayesian literature as a weakly informative prior distribution for ANOVA designs by Rouder et al. (2012) and for logistic regression models by Gelman et al. (2008). Multivariate Cauchy priors have also been proposed by Zellner and Siow (1980) and an extension of these, called the Jeffreys-Zellner-Siow (JZS) prior, has been proposed by Bayarri and García-Donato (2007). The Cauchy distribution with a scale equal to 1, is equivalent to a Student-t distribution with one degree of freedom. A Cauchy distribution centred at zero with a user-specified

[3]The Bayesian Information Criterion (BIC) is a statistical measure used for model selection that balances model fit and complexity by penalising models with more parameters.

Figure 1
Example of a discrete spike and slab prior with a (1) a UI slab component, (2) a Cauchy slab and (3) a Laplace slab component with scales of 0.5, 2.5 and 5.



scale (with a default value of 2.5 following Gelman et al., 2008) is implemented in the R package bgms as the slab component for the discrete spike and slab prior. As can be seen in the middle panel of Figure 1, the Cauchy distribution is characterised by heavier tails relative to the normal distribution, which means that it assigns a relatively higher probability to extreme values compared to the UI prior. Therefore, this prior can capture more extreme values for the edge weights.

3.2.3 Laplace

When used as a univariate shrinkage prior, the Laplace distribution (also known as the double exponential distribution), is called the Bayesian lasso and is well known in the statistical literature (Hans, 2010; Lykou & Ntzoufras, 2013; Narisetty & He, 2014; Park & Casella, 2008; van Erp et al., 2019). Ročková (2018) combines the Bayesian variable selection properties of the spike and slab method with the Laplace shrinkage prior and proposes the spike and slab lasso. Since the widely used frequentist methods for estimating network psychometric models rely heavily on lasso penalization (Friedman et al., 2008; van Borkulo et al., 2014), it is valuable to explore the behavior of the Bayesian lasso when employed in the analysis of MRF models. Therefore, we implemented a Laplace distribution centred at zero with an adjustable scale as the third candidate for the slab component.[4]. By comparing the Cauchy and Laplace distributions in Figure 1, we can see that the Cauchy distribution has slightly heavier tails, which means it assigns a higher probability to extreme values and has more pronounced

[4] In Appendix B, we explain how we implemented this prior in the R package bgms.

tail behavior, while the Laplace is more peaked around zero.

The Jeffreys-Lindley paradox (Lindley, 1957) states that as the scale of the prior distribution increases (indicating greater prior uncertainty), the Bayes factor tends to favor the null hypothesis more strongly. This paradox illustrates that wide ("uninformative") priors can lead to conservative inferences that favor the null hypothesis even when there is evidence in the data against it. The reason is that winder prior distributions lend plausibility to extreme parameter values that the data cannot support; thus, the null hypothesis gains an advantage over the alternative hypothesis. Therefore, we expect that varying the value of the scale of the Cauchy or Laplace slab components will have an effect on the outcome of the inclusion Bayes factor, and we explore the extent of this effect in the simulation study.

3.2.4 Additional Options

Other options for priors on the edge weight parameters that we considered but did not implement in the simulation are: (i) the Horseshoe prior (Carvalho et al., 2009), another popular univariate shrinkage prior, (ii) the g prior (Li & Clyde, 2018), where setting $g = N$ yields the UI prior, (iii) the intrinsic prior (Berger & Pericchi, 1996), (iv) the fractional prior (O'Hagan, 1995), and (v) the penalized complexity prior (Simpson et al., 2017).

3.2.5 Priors in Other R Packages

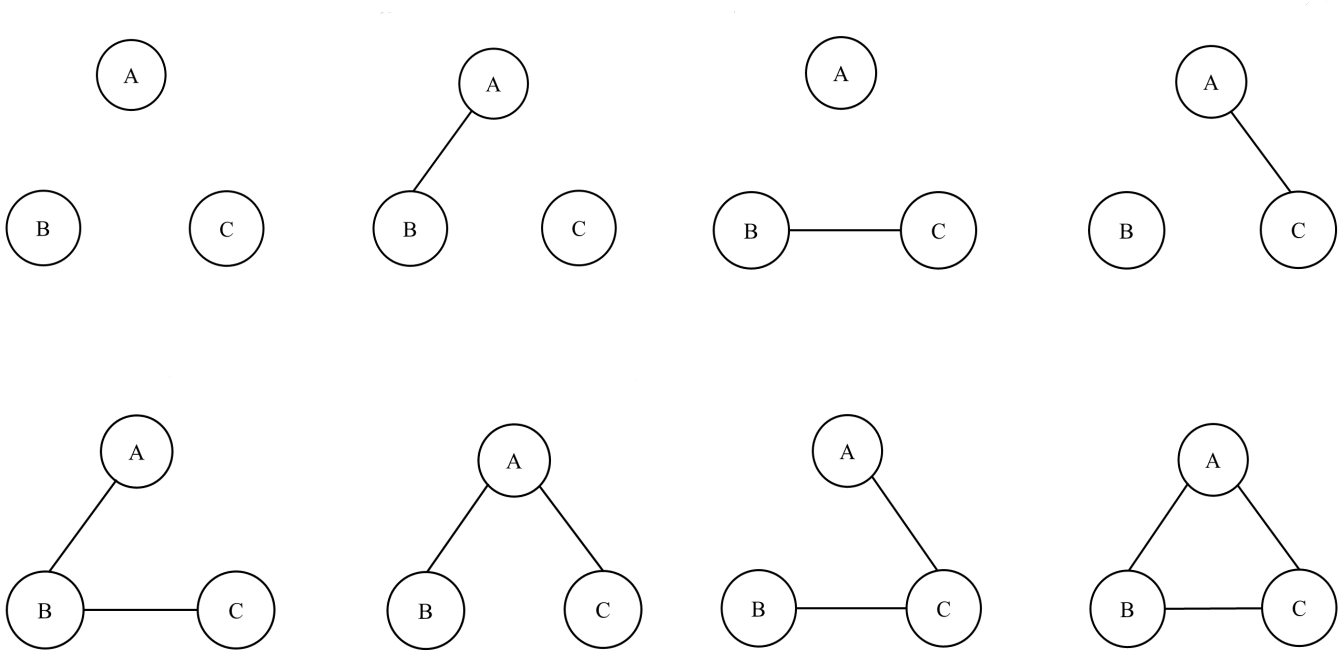
The R package BDgraph (Mohammadi & Wit, 2019) can be used to fit Gaussian graphical models and also uses a BMA approach to account for the structure uncertainty in the estimation of the models. The authors stipulate a G-Wishart(b, D) distribution with $b > 2$ degrees of freedom and D, a scale matrix which is by default a $p \times p$ identity matrix, as a prior on the precision matrix (the inverse of the covariance matrix, which is equivalent to Θ in the ordinal MRF case). The R package BGGM (Williams & Mulder, 2020b), which can also be used for Bayesian estimation of Gaussian graphical models, specifies either a Matrix-F prior or a Wishart distribution on the precision matrix (Williams & Mulder, 2020a). The latter package does not use BMA to account for structure uncertainty but offers a choice of regularization priors (see, Williams & Mulder, 2020b for further details).

3.3 Prior Distributions for the Network Structure — Edge Indicators

To complete the specification of the discrete spike and slab priors, we need to set prior distributions on the edge indicator variables. As we have seen, setting the priors on the edge indicator variables is equivalent to setting the priors on the network (graph) structure (cf. Equations 2 and 4). To help understand these structure priors, consider an example network with three variables: A, B and C. This network can have a maximum of $k = 3(3-1)/2$ possible edges, resulting in $2^3 = 8$ possible structures,

as shown in Figure 2.

Figure 2
All the possible structures for a network of three variables A, B and C.



3.3.1 Bernoulli

In the simplest setup, all of the edge indicator variables are modeled as independent and identically distributed Bernoulli variables,

$$p(\boldsymbol{\gamma} \mid \boldsymbol{\pi}) = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} \pi_{ij}^{\gamma_{ij}} \, (1 - \pi_{ij})^{1-\gamma_{ij}},$$

where π_{ij} denotes the prior edge inclusion probability. In the spirit of 'uninformative' or 'objective' prior distributions, one could choose a special case of this prior by setting $\pi_{ij} = 0.5$ for all γ_{ij}. This particular choice is called the uniform prior, since it assigns equal prior probability to each of the 2^k possible structures (see e.g., Huth et al., 2023; Marsman et al., 2022; Mohammadi & Wit, 2019). For our three variable example in Figure 2, this implies that each structure is given a prior probability of $1/8$; meaning that the structure with no edges, each of the structures with one or two edges, and the fully connected structure will receive the same amount of prior probability. The uniform prior therefore does not take into account the complexity of the structure, that is the number of present edges in a specific structure. A consequence of assigning the same amount of prior probability is that the

uniform prior favours medium sparse structures because they are more numerous relative to sparse and dense structures. For example, in Figure 2, the structures with one or two edges will receive $3 \times {}^{1}/_{8} = {}^{3}/_{8}$ of the prior probability, while the empty structure and the fully connected structure will receive a probability of ${}^{1}/_{8}$.

Researchers may also choose to set the prior inclusion probability to a different value, depending on whether they expect a dense or sparse network structure. For example, if we are quite certain that all three edges in our example should be present, we can set $\pi_{ij} = 0.9$. This would assign a probability of 0.79 to the fully connected structure, a probability of 0.08 to each of the three structures with two edges present, a probability of 0.009 to each of the three structures with one edge present, and a probability of 0.001 to the structure with no edges. Finally, researchers can assign different prior inclusion probabilities for different edges if they have a substantive reasons to do so.

In this paper, we explore the uniform prior because it sets the prior inclusion odds equal to one, resulting in an inclusion Bayes factor equal to the posterior odds (Equation 2). The uniform prior is the default option in the R packages bgms and BDgraph.

3.3.2 Beta-Bernoulli

The uniform prior has been criticized in the BMA literature due to its failure to account for sparsity, collinearity and multiplicity (e.g., Consonni et al., 2018). Scott and Berger (2010) propose to extend the Bernoulli prior by specifying a Beta(α, β) hyperprior on the edge inclusion probability π. This prior is known as the Beta-Bernoulli (or Beta-Binomial) prior and takes the form

$$\begin{aligned} p(\boldsymbol{\gamma}) &= \int_0^1 \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} \theta^{\gamma_{ij}} (1-\theta)^{1-\gamma_{ij}} \frac{1}{\mathsf{B}(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \, \mathsf{d}\theta \\ &= \int_0^1 \theta^{\gamma_{++}} (1-\theta)^{k-\gamma_{++}} \frac{1}{\mathsf{B}(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \, \mathsf{d}\theta \\ &= \frac{\mathsf{B}\left(\alpha + \gamma_{++}, \beta + k - \gamma_{++}\right)}{\mathsf{B}(\alpha, \beta)} \end{aligned}$$

where $\gamma_{++} = \sum_{i=1}^{(p-1)} \sum_{j=i+1}^{p} \gamma_{ij}$, denotes the structure complexity (i.e., the number of present edges in a network). By specifying this prior on the vector of edge indicator variables, we assign the same prior probability to all structure complexities, and then distribute the probability equally within each complexity. For example, in Figure 2 we have four different complexities: (1) one network with no edge; (2) three networks with one edge; (3) three networks with two edges; (4) one fully connected network. The Beta-Bernoulli prior would first assign each of the four complexities a prior probability of ${}^{1}/_{4}$, which will give a prior probability of ${}^{1}/_{4}$ for the no-edge and fully connected networks and a probability of ${}^{1}/_{12}$ for the other networks. This prior is therefore called the hierarchical prior, as it assigns a prior probability to the different structures in a hierarchical fashion. It is also sometimes

referred to as a uniform prior on the structure complexity (Marsman et al., 2022). The default choice of the Beta-Bernoulli prior with $\alpha = \beta = 1$ makes the prior inclusion odds in Equation 2 equal to 1. For example, to compute the prior inclusion odds for an edge between variables A and B, we must first compute the prior inclusion probability by summing the prior probabilities of all structures containing an edge between A and B, and then divide by 1 minus that probability. In this case $p(\mathcal{H}_1^{(AB)} \mid \mathbf{X}) = 3 \times 0.08\bar{3} + 0.25 = 0.5$, so the prior odds are $0.5/0.5 = 1$.

Researchers using the R package bgms are free to set different values for the α and β hyperpriors. For example, if they have good reasons to believe that the network structure should be dense, they should set α larger relative to β. Note, however, that in this case the prior odds are not necessarily equal to one.

3.3.3 Additional Options

In this subsection we briefly mention other options for the prior on the network structure that we have not included in this paper. Wilson et al. (2010) suggests using $\alpha = 1$ and $\beta = \lambda p$ for the Beta-Bernoulli prior; where λ acts as a penalty parameter. For more details on this and other related prior distributions, see Consonni et al. (2018) and references therein. One can also incorporate a random graph model, such as the Stochastic Block Model (SBM, Holland et al., 1983) as a prior on the network structure. This approach allows us to determine whether the nodes belong to one or more clusters. The SBM helps to reveal locally dense structures, meaning that edges within a cluster are more likely to be observed than edges between clusters, allowing researchers to focus on groups of nodes rather than just individual edges between pairs of nodes. This prior can also provide a confirmatory approach to testing a hypothesized clustering structure of the data.

3.4 Prior Distributions for the Category Threshold Parameters

As can be seen from Equation 4, we are also required to specify a prior distribution for the category threshold parameters $p(\mu_{ih})$. Marsman et al. (2023a) specify a generalized Beta-prime $(1/2, 1/2)$ distribution as the prior for the threshold parameters. Overall, the Beta-prime distribution is a versatile tool for modeling positively skewed data with two shape parameters that allow fine-tuning of the behavior of the distribution. Since the category threshold parameters are not of primary interest to researchers, we do not explore further options for their prior distributions and do not vary the hyperparameter values of the Beta-prime distribution in the simulation study.

3.4.1 Options in Other R Packages

Marsman et al. (2022) stipulate independent standard normal distributions for the external field parameters in the Ising model. The R packages BDgraph and BGGM treat the observed (Gaussian) data as centered by setting the mean vector of the multivariate normal distribution of the data to zero, thus eliminating the need to specify prior distributions for the threshold (mean) parameters.

4. SIMULATION DESIGN

Given the inherent complexity of MRF models, it is not feasible to empirically investigate how the priors work. To shed light on this issue, we conducted a simulation study. Within the design, we systematically varied the prior specifications as well as the characteristics of the data, allowing for a rigorous examination of their interplay. We varied: (i) the two priors for the graph structure (the uniform Bernoulli prior and the Beta-Bernoulli prior with $\alpha = \beta = 1$) and (ii) the three densities for the slab component of the discrete spike and slab prior (the UI, Cauchy and Laplace densities). For the Cauchy and Laplace distributions, we additionally varied the scale parameter from 1 to 2.5 with an increment of 0.5. It should be noted that the Cauchy and the Laplace distributions are separate probability distributions with their own properties, and their scales are not related by a fixed factor; therefore, their densities do not match when having the same value for the scale parameter (c.f. Figure 1). For the priors in the (category) threshold parameters, we stipulate a generalized Beta-prime ($1/2, 1/2$) distribution.

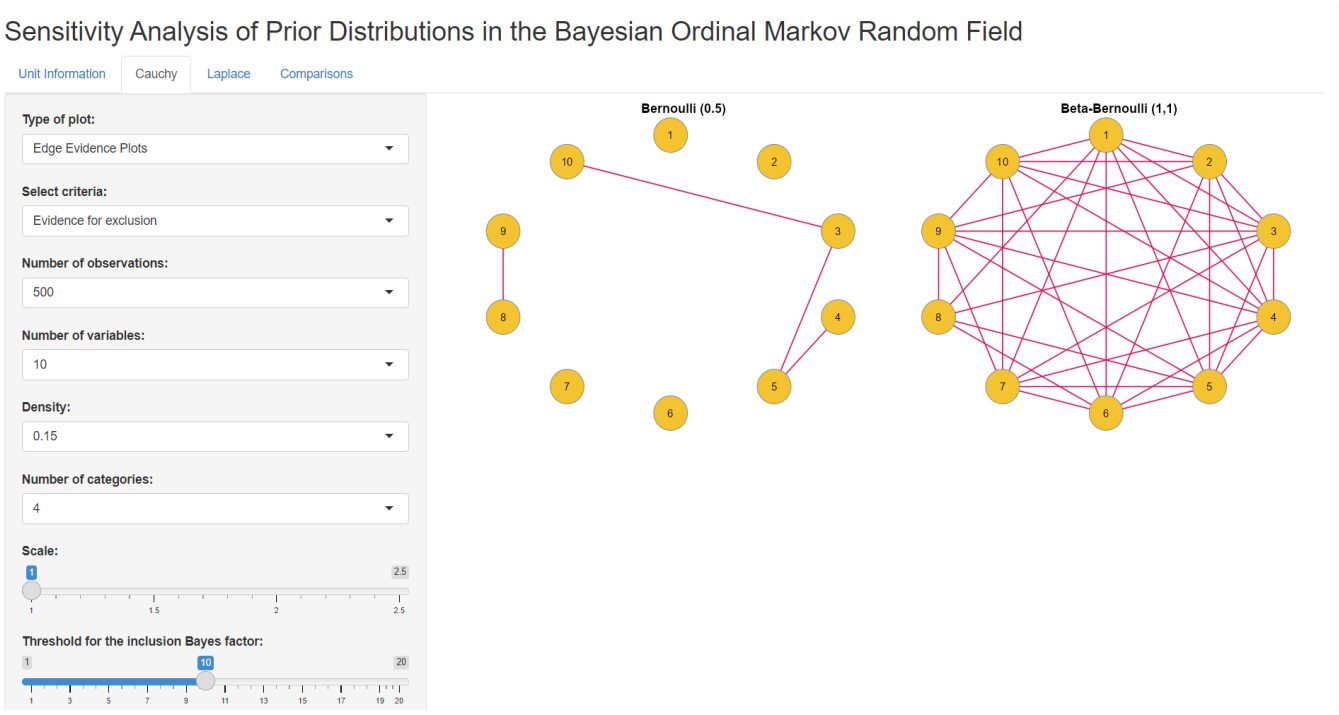
For the data, we varied: (i) the sample size $N \in \{100, 500, 1000\}$; (ii) the number of variables in the network $p \in \{5, 10, 15\}$; (iii) the proportion of present edges in the network $d \in \{0.15, 0.5, 0.95\}$, corresponding to a sparse, medium and dense network respectively; and (iv) the number of ordinal categories $c \in \{2, 4\}$, corresponding to an Ising (binary) and four-category ordinal model, respectively. Note that the number of categories is not required to be the same for all variables; however, for simplicity, we keep them fixed in this simulation. We simulated 100 data sets for each of the 54 combinations to account for the random fluctuations of individual data sets and approximate the population outcomes.

To generate data based on the aforementioned characteristics we took the following steps. First, we generated data sets with a sample size of 1000 from a unidimensional Item Response Theory (IRT) model for ordinal data known as the Generalized Partial Credit Model (GPCM, Muraki, 1990); which ensures that the underlying network structures are fully connected with sufficiently large edge weight parameters. For a discussion of how IRT models relate to network models, see Marsman et al. (2018), with a focus on the relation between the Ising model and the Rasch model. We then obtained maximum pseudolikelihood estimates on the IRT data sets separately for all three network density levels d, to account for the fact that the strength of the remaining edge weight changes as previously existing edges are removed; this is known as the boundary specification problem (e.g., Laumann et

al., 1989; Neal & Neal, 2023). We induced sparsity by giving each edge a probability of inclusion of 0.95, 0.5, and 0.15 for the dense, medium, and sparse networks, respectively.

For each of the 54×100 datasets, we analyzed the corresponding MRF model using a version of the R package `bgms` modified for the purpose of this study.[5] In total, this setup yielded 972 simulation conditions with 972×100 models. We used the Dutch national supercomputer Snellius to perform the analyses and ran each model using $10{,}000$ iterations.[6] In situations where we obtained a posterior edge inclusion probability of 1, the inclusion Bayes factor was undefined (cf. Equation 2); therefore we set these values to $\text{BF}_{10} = 100$. The code for this simulation is openly available at https://osf.io/3mqgc/.

Figure 3
Screenshot of the Shiny app setup.



Given the complexity of our simulation design, we have chosen to present the results in two ways. In the following section, we present the main conclusions of the simulation by exploring the most important patterns. In addition, to take full advantage of the size of the simulation, we provide an interactive Shiny app (Chang et al., 2022); that makes it easy to navigate and compare the results across different simulation conditions. This will help researchers to explore specific settings relevant to their research interests. The app can be accessed at https://nikolasekulovski.shinyapps.io/bgm_priors/. In the Shiny app, the results for each of the slab components are presented in three separate tabs, each containing two plots for the two different priors on the structures (Figure 3). A fourth tab is available where researchers can compare the performance of pairs of slab components within

[5] https://github.com/sekulovskin/bgms_V2
[6] https://www.surf.nl/en/dutch-national-supercomputer-snellius

each structure prior. On the left side, the user can navigate and change the simulation settings. Two types of plots are available: (i) a plot of the posterior mean estimates for the edge-weight parameters against their respective inclusion Bayes factors; (ii) edge evidence plots showing which edges have evidence for inclusion, exclusion, and inconclusive evidence.

5. RESULTS

In this section, we briefly summarize the results of the simulation study by highlighting some of the key findings. We make use of boxplots as an efficient way to summarize the large number of simulation settings. A boxplot is a graphical representation of numerical data by showing the median, interquartile range, and outliers, providing a concise summary of the distribution and variability of the log inclusion Bayes factors. We present the results by contrasting the two structure priors for each slab component. We also pay special attention to the effect of scale for the Cauchy and Laplace slab components, for which an additional analysis is performed. In this paper, we compare the relative distributions of the log inclusion Bayes factors as continuous measures of evidence. Thus, even though we generally consider the cutoffs of 10 and $1/10$ (e.g., Huth et al., 2023), we are still talking about the relative strengths of the Bayes factor as a continuous measure of evidence. Thus, we do not use performance measures such as AUC values (e.g., Fawcett, 2006).

5.1 Unit Information

Figure 4 shows that the evidence for both edge exclusion and edge inclusion gets stronger as the sample size increases. Moreover, for the ordinal data with 4 categories, we observe stronger evidence for edge inclusion compared to the binary data; it is obvious from Equation 1 that the term $X_i X_j$ can grow larger as a function of m, which leads to this result. Furthermore, across all simulation settings, the evidence for edge exclusion is stronger under the Beta-Bernoulli prior (observe for example, the last row of Figure 4 for $p = 15$). Of course, this is most obvious for the low-density structures (i.e., when most edges are absent). The Beta-Bernoulli prior takes into account the complexity of the structure and can therefore uncover evidence for exclusion faster than the (uniform) Bernoulli prior. The discrepancy between the two structure priors becomes more obvious as the number of nodes p in the network decreases. In terms of evidence for edge inclusion, however, we do not observe much difference between the two structure priors. The interested reader can inspect the "Bayes Factor vs. Edge Weights" plot in the Shiny app to further explore these results. Finally, a very striking observation that is consistent across all three slab components (see also Figure 5 and 6) is the tail behavior of the inclusion Bayes factor when the true network structure is sparse (yellow boxplots). The edge weight parameters in the ordinal MRF are not standardized; this means that their range depends on the size of the network and the number of response categories. By inducing sparsity and decreasing the number of variables or increasing the number of response categories, the absolute value of

the edge weight parameters tends to increase; this is, in part, related to the boundary specification problem we mentioned earlier (cf. Laumann et al., 1989; Neal & Neal, 2023). For the present edges in a sparse network, the corresponding edge weight parameters are larger relative to their counterparts in the dense networks (of the same size); this, in turn, increases the posterior inclusion probability yielding a larger inclusion Bayes factor.

Figure 4

Boxplots showing the log inclusion Bayes factors for the setting with a unit information slab component. The x-axis denotes the different sample sizes (N) and the different rows denote the different number of variables (p). Each row contains two panels; (left) for the Bernoulli structure prior and (right) for the Beta Bernoulli structure prior. Each panel is split in two, with the first part showing the log Bayes computed for the binary data (2 categories) and the second for the ordinal data (4 categories). The color scheme represents network density, with yellow (left) indicating sparse, blue (middle) indicating medium, and purple (right) indicating dense networks. The dashed horizontal lines denote cutoff values for log Bayes factors of 2.3, 0 and -2.3, indicating evidence of edge inclusion, completely inconclusive evidence and exclusion, respectively.

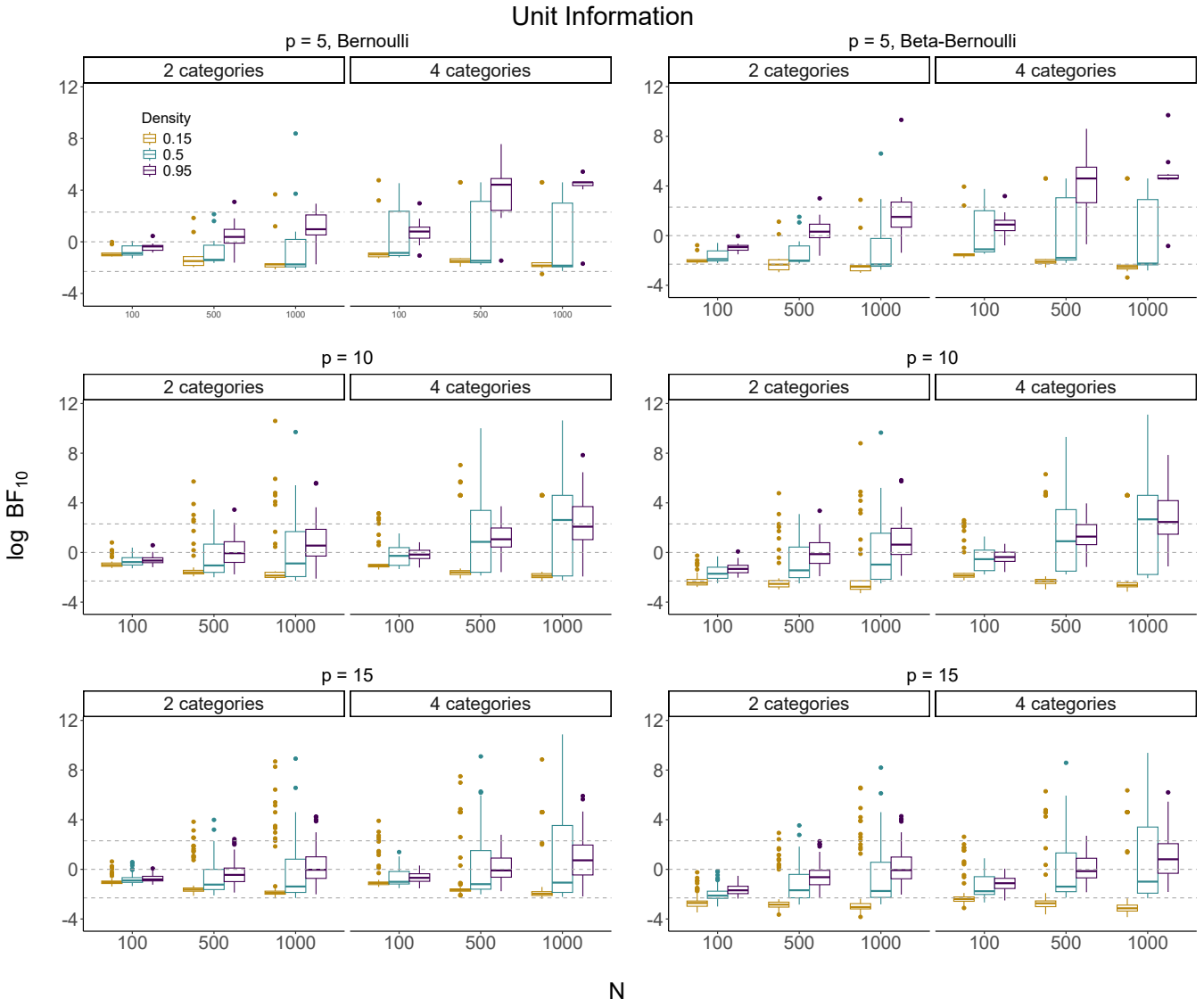
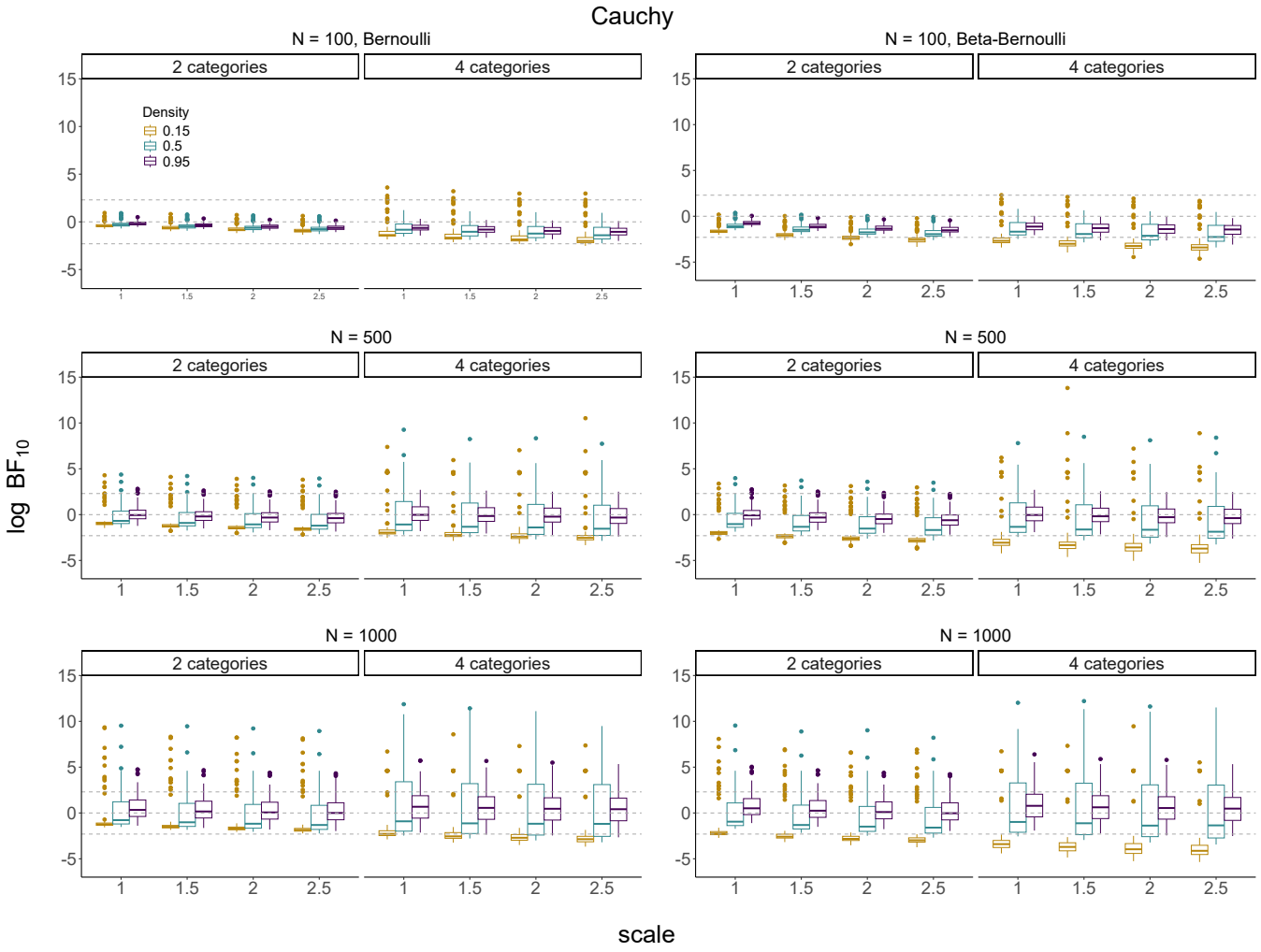


Figure 5

Boxplots showing the log inclusion Bayes factors for the setting with a Cauchy slab component for $p = 15$ variables. The x-axis denotes the different values for the scale of the slab and the different rows denote the sample size (N). Each row contains two panels; (left) for the Bernoulli structure prior and (right) for the Beta Bernoulli structure prior. Each panel is split in two, with the first part showing the log Bayes computed for the binary data (2 categories) and the second for the ordinal data (4 categories). The color scheme represents network density, with yellow (left) indicating sparse, blue (middle) indicating medium, and purple (right) indicating dense networks. The dashed horizontal lines denote cutoff values for log Bayes factors of 2.3, 0 and -2.3, indicating evidence of edge inclusion, completely inconclusive evidence and exclusion, respectively.



5.2 Cauchy

Due to the large simulation design for the Cauchy and Laplace priors, we only show the results for the case of $p = 15$ variables, since, as we saw in the previous subsection, the patterns are most pronounced with many variables and generalize to the situation with fewer variables. The interested reader can explore the conditions with 5 and 10 variables in the Shiny app. Figure 5 shows the log inclusion Bayes factors under the Cauchy slab component plotted as a function of scale, for the Bernoulli and Beta-Bernoulli structure priors, respectively. It is immediately apparent that the evidence in favor

of edge exclusion (i.e., in favor of the null hypothesis) increases as the value of the scale increases (see, for example, the lower left panel of Figure 5). The interested reader is referred to the "Edge Evidence Plots" in the Shiny app to visually verify that there is a larger difference between the two structure priors; and that the evidence for exclusion is more strongly affected than the evidence for inclusion. As with the UI slab, the evidence in favor of edge exclusion is greater under the Beta-Bernoulli structure prior. The results are stronger for larger sample sizes for the four-category ordinal data.

5.3 Laplace

In Figure 6 we observe an almost identical pattern as for the Cauchy slab. A notable difference is that the results under the Laplace slab, especially in terms of evidence for exclusion, are slightly less strong than under the Cauchy prior. The reader can explore this further in the "Comparisons" tab of the Shiny app. We now turn to the effect of the scale of the slab component.

5.4 The Scale of the Slab Component

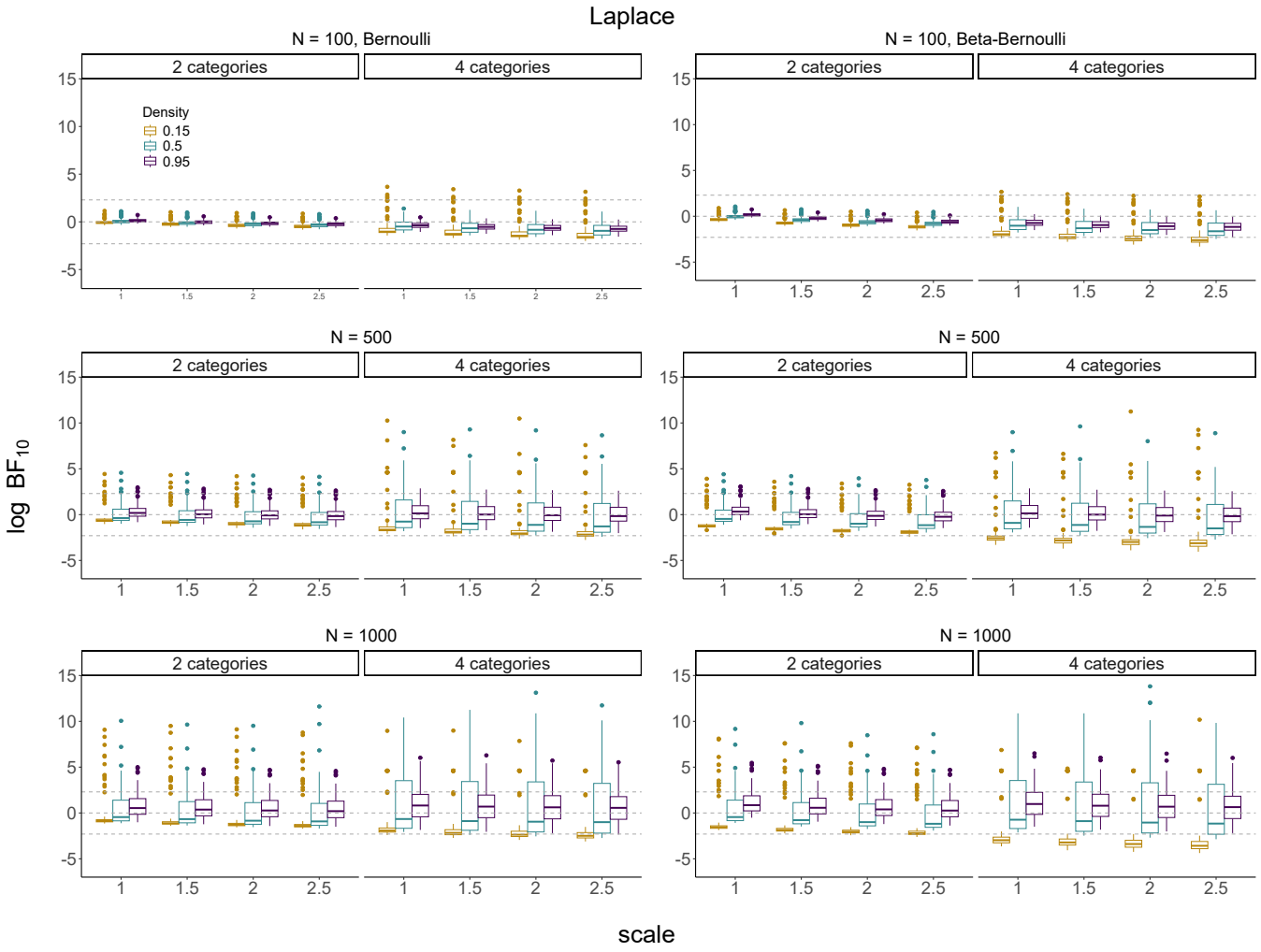
One insight of the Jeffreys-Lindley paradox (Lindley, 1957) is that as the complexity of a hypothesis increases (in this case, the complexity of $\mathcal{H}_1^{(ij)}$ increases with the scale of the slab component), the Bayes factor begins to favor the simpler hypothesis (i.e., $\mathcal{H}_0^{(ij)}$), even if $\mathcal{H}_1^{(ij)}$ happens to be the true hypothesis. This seemingly paradoxical result is due to the Bayes factor's innate tendency to penalize for model complexity, acting as an Occam's razor (see McFadden, 2023 for a more general discussion of how the Bayesian approach naturally incorporates Occam's razor). Consequently, in such cases, the evidence in the data against the null hypothesis is not strong enough to offset the complexity penalty, resulting in a Bayes factor that favors the null hypothesis. As we have already seen in the results of the main simulation, this is also the case for the inclusion Bayes factor. Furthermore, based on the main results, we have reason to suspect that due to the inherent complexity of MRF models, additional factors such as the size and density of the network may play a confounding role on the effect of the size of the slab component on the inclusion Bayes factor.

To further investigate the effect of the scale of the slab component, we simulated 100 data sets, each with $p \in \{10, 11, 12, 13, 14, 15\}$ and $N = 500$ observations in turn for $d \in \{1, 0.5, 0\}$, i.e., from a fully dense, a moderately dense, and a fully sparse graph structure. We analyzed the models using a Cauchy slab component with scale values ranging from 1 to 2.5 with an increment of 0.5. We stipulated a Bernoulli prior on the network structure.

Figure 7 illustrates that as the scale of the slab component increases, $\mathcal{H}_0^{(ij)}$ gains slightly more support, even though $\mathcal{H}_1^{(ij)}$ is the true hypothesis (one can observe that the box plots and median lines shift downward as the value of the scale increases), regardless of the size and density of the network; as

Figure 6

Boxplots showing the log inclusion Bayes factors for the setting with a Laplace slab component for $p = 15$ variables. The x-axis denotes the different values for the scale of the slab and the different rows denote the sample size (N). Each row contains two panels; (left) for the Bernoulli structure prior and (right) for the Beta Bernoulli structure prior. Each panel is split in two, with the first part showing the log Bayes computed for the binary data (2 categories) and the second for the ordinal data (4 categories). The color scheme represents network density, with yellow (left) indicating sparse, blue (middle) indicating medium, and purple (right) indicating dense networks. The dashed horizontal lines denote cutoff values for log Bayes factors of 2.3, 0 and -2.3, indicating evidence of edge inclusion, completely inconclusive evidence and exclusion, respectively.

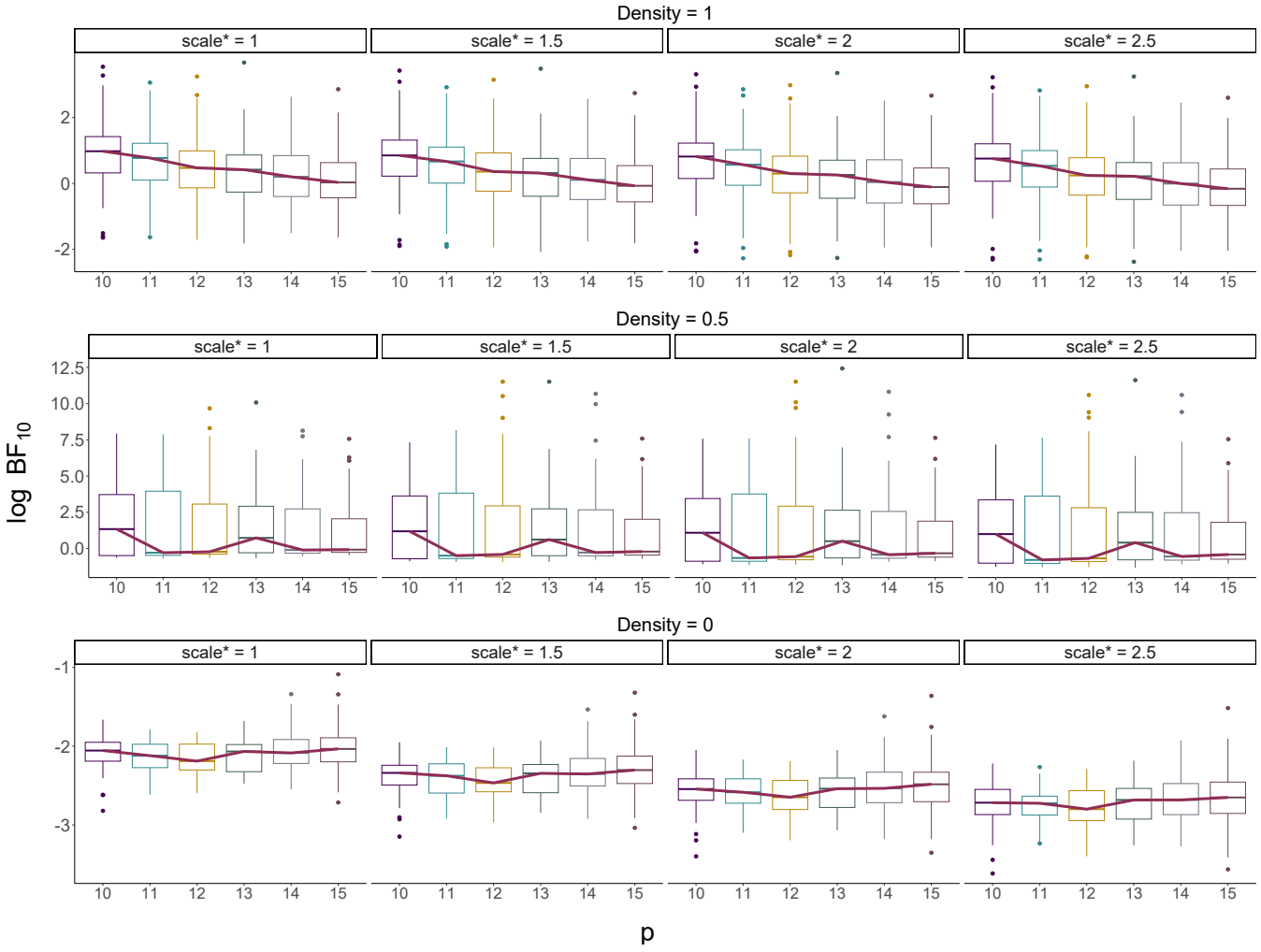


we discussed earlier, this is a well-known result in Bayesian statistics and is consistent across all three network density levels (cf. Figures 5 and 6). The more interesting problem, specific to the inclusion Bayes factor, arises when we consider the joint effect of network size and density.

In the first row of Figure 7, when all edges are present (i.e., the network is generated from a dense graph), we observe that the inclusion Bayes factor decreases as the number of variables in the network increases (as seen by the downward slope of the median lines). This is again due to the fact that the corresponding edge weight parameters shrink as the network size increases, and is related to the

Figure 7

Boxplots showing the log inclusion Bayes factors as a function of the number of variables p (the boxplots are also colored with respect to p, for easier differentiation). Different columns represent the 4 scale values for the Cauchy slab component, ordered from the lowest "scale = 1" to the highest "scale = 2.5". The first row shows the results for fully connected structures (Density = 1), the second row for medium-dense structures (Density = 0.5), and the last row for sparse structures (Density = 0). The solid lines represent the relative change of the median inclusion of Bayes factors over p.



fact that the edge weight parameters are not standardized. Since the pattern is consistent, a correction may be possible by decreasing the scale of the slab component as a function of the number of variables p. Such an option is presented in Appendix C.

In the middle row of Figure 7, when half of the edges are present (i.e., a medium-dense network), we do not observe a consistent pattern with respect to the number of variables as we did in the previous case (as observed by the median lines). In such situations, it becomes more difficult to find a one-size-fits-all solution and scale the slab component as a function of the number of variables p; but it also becomes less necessary to do so, since we do not observe a consistent bias, as we did in the dense situation.

The last row of figure 7 shows the situation when the network is completely sparse, here there is not much difference between the spread of the Bayes factor for different values of p within the same scale (as seen by the almost flat median lines). Since (almost) all edges are missing, the effect of the number of variables does not affect the inclusion Bayes factor. This is because all edge weight parameters are equal or close to zero and the null hypothesis is true.

6. STREAMLINING SIMULATION STUDIES WITH THE R PACKAGE SIMBGMS

As we hope to have demonstrated in this paper, simulation studies serve as a valuable tool for assessing the performance and validity of statistical methods, especially for complex models such as MRFs. Such studies not only validate the theoretical underpinnings, but also provide insight into the practical applicability of the methods in different scenarios. Thus, the benefits of simulation studies should also be used by applied researchers to inform their choices for the analysis of empirical data using network psychometric models (Hoekstra et al., 2023). The simulations can be performed in the pre-registration process to help the researcher decide on sample size, number of variables to include, and analysis choices such as setting priors. In practice, however, simulation studies are typically used only by researchers with strong methodological and programming backgrounds. In this context, we introduce the new user-friendly simBgms package, which aims to streamline the complicated process of running simulation studies for Bayesian MRF models.

The primary aim of the simBgms package is to simplify and accelerate the execution of simulation studies for Bayesian graphical models. By encapsulating complex procedures in an easy-to-use interface, simBgms allows researchers to efficiently generate and analyze data from MRF models (such as Ordinal MRF and the Gaussian graphical model) using the R packages bgms (Marsman et al., 2023b) and BDgraph (Mohammadi & Wit, 2019). The package seamlessly integrates data simulation and model estimation to provide a holistic framework for conducting simulation studies. Notable features include the ability to parallelize model estimation (especially useful when running analyses on supercomputers, i.e., clustered computers), which improves computational efficiency, the flexibility to specify simulation parameters tailored to specific research contexts, as well as providing summarized (ready-to-plot) results.

6.1 Installation & Usage

The simBgms package is currently available on GitHub.[7] By using the packages devtools (Wickham et al., 2022) and remotes (Csárdi et al., 2023), users can easily install the package as shown in the code below.

[7] Once fully developed, the authors aim to make the package available on The Comprehensive R Archive Network. For now, the package is available at https://github.com/sekulovskin/simBgms.

```
if (!requireNamespace("remotes")) {
  install.packages("remotes")
}
devtools::install_github("sekulovskin/simBgms")
```

Once installed, researchers can use the main sim_bgm function to initiate simulation studies. Below is an example that specifies the same design as the main simulation presented in this paper, with a Cauchy slab component and a Bernoulli structure prior. Note that only the Cauchy slab component is available in the main version of the bgms package.

```
library(simBgms)

sim_bgm(level = "Discrete",
        variable_type = "ordinal",
        repetitions = 100,
        no_observations = c(100, 500, 1000),
        no_variables = c(5, 10, 15),
        no_categories = c(2, 4),
        induce_sparsity = TRUE,
        density = c(0.15, 0.5, 0.95),
        irt = TRUE,
        interaction_scale = seq(1, 2.5, 0.5),
        edge_prior = "Bernoulli",
        iter = 1e4,
        burnin = 1e3)
```

For the ordinal and binary data, the parameters used to generate the synthetic data can be obtained in two ways, either from an IRT model using the irt = TRUE argument, or by the user providing a custom dataset on which the parameters are estimated to generate the synthetic data. See the package documentation for detailed usage instructions and additional examples. In addition, the separate functions generate_data and estimate_models provide the flexibility to generate data and estimate models independently, allowing for custom analyses as needed.

The simBgms package is currently in a development stage, it promises further refinement and expansion in future versions. Ongoing efforts are focused on incorporating additional functionality, especially with respect to flexibility in data generating process.

7. DISCUSSION

The goal of this paper was to provide an accessible and extensive exploration of the prior distributions for network structure and edge weights in Markov random field models for ordinal and binary data. We illustrated the behavior of these priors when used to test for conditional independence (edge absence) through a large simulation study. We also made the extensive simulation results more accessible by also presenting them in a Shiny app. Finally, we introduced a user-friendly R package that makes the benefits of running simulation studies for Bayesian analysis of MRF models for binary, ordinal, and continuous data more accessible to applied researchers. In the remainder of this section, we briefly comment on the prior robustness of the ordinal MRF, discuss the limitations of the current work, and provide directions for future research.

7.1 Prior Robustness

The Beta-Bernoulli prior consistently yields stronger evidence for edge exclusion than the Bernoulli prior, highlighting the importance of considering the network complexity when specifying priors on the network structure. Furthermore, the results indicate that both sets of priors have a stronger influence on the evidence for edge exclusion than for edge inclusion; this is somewhat more pronounced for the Cauchy prior than for the unit information and Laplace priors. In general, the behavior of the three slab components is relatively similar with respect to their interaction with the structure priors and the data characteristics. However, the scale of the slab component has a significant effect on the inclusion Bayes factor; moreover, for dense graphs, the inclusion Bayes factor decreases as the number of variables increases, even when the value of the scale is kept fixed. Therefore, in Appendix C, we discuss the possibility of shrinking the scale as a function of the number of variables. Although in practice the number of variables in the network is fixed, it is important to provide a preliminary solution to this theoretical problem. In addition, there may be situations where researchers would like to make the results of networks of different sizes comparable, for example, when comparing data that they expect to behave similarly but are of different sizes. We have shown that the fact that the edge weight parameters are not standardized explains some of the effects we saw, especially with respect to the size of the scale of the slab component and some of the discrepancy between ordinal and binary data. Finally, we have seen that in most cases, holding the prior specification fixed, the inclusion Bayes factors computed on the ordinal data show stronger evidence in favor of the truth compared to the binary data.

7.2 Limitations & Future Directions

While our simulation study provides valuable insights, there are limitations to consider; we explored a number of scenarios but did not exhaust all possible conditions. For example, we kept the number

of categories the same across all variables for the ordinal data. In addition, our recommendations are based on binary and ordinal data with four categories, which limits the conclusions we can draw. Also, we explored only two specific settings of the Bernoulli and Beta-Bernoulli priors, so we did not illustrate how the results of the Bayes factor change when we deviate from these default choices. We only touched on the relationship between the size of the slab component and the number of variables in the network, providing a generic solution without exploring the underlying reasons why such a solution works. We also did not explore the interplay between the size of the slab component and the number of ordinal categories.

Based on the current results, future research should focus on standardizing and/or scaling the edge weight parameters through the prior distribution in the estimation process or by adjusting the model parameters directly through the likelihood function. For the former, multivariate prior distributions (such as a multivariate Cauchy distribution or a Wishart distribution) could be considered as possible candidates for the slab component, as they can adjust the values of the parameters with respect to the size of the network.

7.3 Concluding Remarks

We hope that the simulation study, combined with the theoretical insights, will provide guidance to researchers navigating the complex terrain of prior selection in network psychometrics. By considering the network complexity, data characteristics, and sample size, researchers can make informed choices that are consistent with the specific goals of their analyses. The guidelines and tools provided in this paper are intended to enable researchers to make effective use of Bayesian model averaging and to increase the robustness of their results in applied research settings. As the Bayesian approach to network psychometrics continues to evolve, we look forward to extending these insights to mixed data types and introducing substantively motivated prior choices for the network structure and edge weight parameters.

7.4 Funding

NS, SK, and MM were supported by the European Union (ERC, BAYESIAN P-NETS, #101040876). JMBH has been supported by the gravitation project 'New Science of Mental Disorders' (www.nsmd.eu), supported by the Dutch Research Council and the Dutch Ministry of Education, Culture and Science (NWO gravitation grant number 024.004.016). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

7.5 Conflict of Interests

The authors declare that they have no conflicts of interest regarding the publication of this paper.

7.6 Reproducibility Statement

The data and materials for all simulation examples are available at https://osf.io/3mqgc/.

REFERENCES

- Anderson, C. J., & Vermunt, J. K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*(1), 81–121. https://doi.org/10.1111/0081-1750.00076
- Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics*, *32*(3), 870–897. https://doi.org/10.1214/009053604000000238
- Bayarri, M. J., Berger, J. O., Forte, A., & García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, *40*(3), 1550–1577. https://doi.org/10.1214/12-AOS1013
- Bayarri, M. J., & García-Donato, G. (2007). Extending conventional priors for testing general hypotheses in linear models. *Biometrika*, *94*(1), 135–152. https://doi.org/10.1093/biomet/asm014
- Berger, J. O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, *25*(3), 303–328. https://doi.org/10.1016/0378-3758(90)90079-A
- Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*(433). https://doi.org/10.1080/01621459.1996.10476668
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2), 192–236.
- Borsboom, D. (2022). Possible futures for network psychometrics. *Psychometrika*, *87*(1), 253–265. https://doi.org/10.1007/s11336-022-09851-z
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., Borkulo, C. D. van, Bork, R. van, & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 58. https://doi.org/10.1038/s43586-021-00055-w
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. *Proceedings of the twelfth International Conference on Artificial Intelligence and Statistics*, 73–80.
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2022). *Shiny: Web application framework for R* [R package version 1.7.4].
- Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034
- Consonni, G., Fouskakis, D., Liseo, B., & Ntzoufras, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, *13*(2), 627–679. https://doi.org/10.1214/18-BA1103
- Cox, D. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series B (Applied Statistics)*, *21*(2), 113–120. https://doi.org/10.2307/2346482

- Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan, M., & Tenenbaum, D. (2023). *Remotes: R package installation from remote repositories, including 'github'* [R package version 2.4.2.1]. https://CRAN.R-project.org/package=remotes
- Dablander, F., & Hinne, M. (2019). Node centrality measures are a poor substitute for causal inference. *Scientific Reports*, *9*(1), 6846. https://doi.org/10.1038/s41598-019-43033-9
- Eicher, T., Papageogiou, C., & Raftery, A. (2007). Determining growth determinants: Default priors and predictive performance in bayesian model averaging. *Journal of Applied Econometrics*, *26*, 30–55. https://doi.org/10.1002/jae.1112
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167
- Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One*, *12*(e0179891). https://doi.org/10.1371/journal.pone.0179891
- Epskamp, S., Waldorp, L., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823
- Etz, A., & Wagenmakers, E.-.-J. (2017). J. B. S. Haldane's contribution to the Bayes factor hypothesis test. *Statistical Science*, *32*, 313–329. https://doi.org/10.1214/16-STS599
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
- Fernandez, C., Ley, E., & Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*(2). https://doi.org/10.1016/S0304-4076(00)00076-2
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191
- George, E. I. (1999). Discussion of "Bayesian model averaging and model search strategies" by Clyde M. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian statistics* (pp. 175–177, Vol. 6). Oxford University Press.
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–889. https://doi.org/10.2307/2290777
- Hans, C. (2010). Model uncertainty and variable selection in Bayesian lasso regression. *Statistics and Computing*, *20*, 221–229. https://doi.org/10.1007/s11222-009-9160-9
- Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215. https://doi.org/10.1177/251524591989865
- Hoekstra, R. H., Epskamp, S., Finnemann, A. T., & Borsboom, D. (2023). Unlocking the poten-

tial of simulation studies in network psychometrics: A tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/4j3hf

- Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–401.
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137. https://doi.org/10.1016/0378-8733(83)90021-7
- Huth, K., de Ron, J., Goudriaan, A. E., Luigjes, K., Mohammadi, R., van Holst, R. J., Wagenmakers, E., & Marsman, M. (2023). Bayesian analysis of cross-sectional networks: A tutorial in R and JASP. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/25152459231193334
- Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258. https://doi.org/10.1007/BF02980577
- Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, *31*, 203–222.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Johal, S. K., & Rhemtulla, M. (2021). Comparing estimation methods for psychometric networks with ordinal data. *Psychological Methods*. https://doi.org/10.1037/met0000449
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091
- Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relation to the Schwarz criterion. *Journal of the American Statistical Association*, *90*(431), 928–934. https://doi.org/10.1080/01621459.1995.10476592
- Keetelaar, S., Sekulovski, N., Borsboom, D., & Marsman, M. (2024). Comparing maximum likelihood and maximum pseudolikelihood estimators for the ising model. *Advances in Psychology*, *2*, e25745. https://doi.org/10.56296/aip00013
- Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications* (Vol. 1). American Mathematical Society.
- Laumann, E. O., Marsden, P. V., & Prensky, D. (1989). The boundary specification problem in network analysis. In L. C. Freeman, D. R. White, & A. K. Romney (Eds.), *Research Methods in Social Network Analysis*. George Mason University Press.
- Lauritzen, S. (2004). *Graphical models*. Oxford University Press.
- Ley, E., & Steel, M. F. J. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, *24*(4), 651–674. https://doi.org/10.1002/jae.1057
- Li, Y., & Clyde, M. A. (2018). Mixtures of g-priors in generalized linear models. *Journal of the American Statistical Association*, *113*(524), 1828–1845. https://doi.org/10.1080/01621459.2018.1469992
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

- Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*(1/2), 187–192. https://doi.org/10.2307/2333251
- Ly, A., Marsman, M., Verhagen, J., Grasman, R. P. P. P., & Wagenmakers, E. J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, *80*, 40–55. https://doi.org/10.1016/j.jmp.2017.05.006
- Ly, A., & Wagenmakers, E.-J. (2021). Bayes factors for peri-null hypotheses. *Test*, *31*(4), 1121–1142. https://doi.org/10.1007/s11749-022-00819-w
- Lykou, A., & Ntzoufras, I. (2013). On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, *23*, 361–390. https://doi.org/10.1007/s11222-012-9316-x
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., Bork, R. v. van, Waldorp, L., Maas, H. v. d., & Maris, G. (2018). An introduction to network psychometrics: Relating ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35. https://doi.org/10.1080/00273171.2017.1379379
- Marsman, M., Huth, K., Waldorp, L. J., & Ntzoufras, I. (2022). Objective Bayesian edge screening and structure selection for Ising networks. *Psychometrika*, *87*(1), 47–82. https://doi.org/10.1007/s11336-022-09848-8
- Marsman, M., Maris, G. K. J., Bechger, T. M., & Glas, C. A. W. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(9050). https://doi.org/10.1038/srep09050
- Marsman, M., & Rhemtulla, M. (2022). Guest editors' introduction to the special issue "network psychometrics in action": Methodological innovations inspired by empirical problems. *Psychometrika*, *87*(1), 1–11. https://doi.org/10.1007/s11336-022-09861-x
- Marsman, M., van den Bergh, D., & Haslbeck, J. M. B. (2023a). Bayesian analysis of the ordinal Markov random field. *PsyArXiv*. https://doi.org/10.31234/osf.io/ukwrf
- Marsman, M., Huth, K., Sekulovski, N., & van den Bergh, D. (2023b). *Bgms: Bayesian variable selection for networks of binary and/or ordinal variables* [R package version 0.1.1]. https://CRAN.R-project.org/package=bgms
- McFadden, J. (2023). Razor sharp: The role of Occam's razor in science [Epub 2023 Nov 29]. *Ann N Y Acad Sci*, *1530*(1), 8–17. https://doi.org/10.1111/nyas.15086
- Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032. https://doi.org/10.1080/01621459.1988.10478694
- Mohammadi, A., & Wit, E. C. (2015). Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis*, *10*(1), 109–138. https://doi.org/10.1214/14-BA889
- Mohammadi, R., & Wit, E. C. (2019). BDgraph: An R package for Bayesian structure learning in graphical models. *Journal of Statistical Software*, *89*(3). https://doi.org/10.18637/jss.v089.i03
- Muraki, E. (1990). Fitting a polytomous item response model to likert-type data. *Applied Psychological Measurement*, *14*(1), 59–71. https://doi.org/10.1177/014662169001400106
- Narisetty, N. N., & He, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *The Annals of Statistics*, *42*(2), 789–817. https://doi.org/10.1214/14-AOS1207

- Neal, Z. P., & Neal, J. W. (2023). Out of bounds? The boundary specification problem for centrality in psychological networks. *Psychological Methods*, *28*(1), 179–188. https://doi.org/10.1037/met0000426
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 99–118. https://doi.org/10.1111/j.2517-6161.1995.tb02017.x
- Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337
- Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404
- Ročková, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *The Annals of Statistics*, *46*(1), 401–437. https://doi.org/10.1214/17-AOS1554
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of mathematical psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001
- Rozanov, Y. A. (1982). *Markov random fields*. Springer-Verlag.
- Ryan, O., Bringmann, L. F., & Schuurman, N. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(6), 953–970. https://doi.org/10.1080/10705511.2022.2056039
- Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Annals of Statistics*, *38*(5), 2587–2619. https://doi.org/10.1214/10-AOS792
- Sekulovski, N., Keetelaar, S., Huth, K., Wagenmakers, E.-J., Bork, R. van, van den Bergh, D., & Marsman, M. (2024). Testing conditional independence in psychometric networks: An analysis of three bayesian methods. *Multivariate Behavioral Research*, 1–21. https://doi.org/10.1080/00273171.2024.2345915
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sorbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-STS576
- Suggala, A. S., Yang, E., & Ravikumar, P. (2017). Ordinal graphical models: A tale of two approaches. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning* (pp. 3260–3269, Vol. 70). PMLR.
- Tadesse, M. G., & Vanucci, M. (Eds.). (2022). *Handbook of Bayesian Variable Selection*. CRC Press.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
- van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scien-*

tific Reports, *4*(5918). https://doi.org/10.1038/srep05918

- van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, *28*, 31–50. https://doi.org/10.1016/j.jmp.2018.12.004
- Vanucci, M. (2022). Handbook of bayesian variable selection. In M. G. Tadesse & M. Vanucci (Eds.). CRC Press.
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulleting & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1343-3
- Wickham, H., Hester, J., Chang, W., & Bryan, J. (2022). *devtools: Tools to Make Developing R Packages Easier* [R package version 2.4.5]. https://CRAN.R-project.org/package=devtools
- Williams, D. R. (2021a). Bayesian estimation for Gaussian graphical models: Structure learning, predictability, and network comparisons. *Multivariate Behavioral Research*, *56*(2), 336–352. https://doi.org/10.1080/00273171.2021.1894412
- Williams, D. R. (2021b). The confidence interval that wasn't: Bootstrapped "confidence intervals" in L_1-regularized partial correlation networks. *PsyArXiv*. https://doi.org/10.31234/osf.io/kjh2f
- Williams, D. R., & Mulder, J. (2020a). Bayesian hypothesis testing for Gaussian graphical models: Conditional independence and order constraints. *Journal of Mathematical Psychology*, *99*(102441). https://doi.org/10.1016/j.jmp.2020.102441
- Williams, D. R., & Mulder, J. (2020b). BGGM: Bayesian Gaussian graphical models in R. *Journal of Open Source Software*, *5*(51), 2111. https://doi.org/10.21105/joss.02111
- Wilson, M. A., Iversen, E. S., Clyde, M. A., Schmidler, S. C., & Schildkraut, J. M. (2010). Bayesian model search and multilevel inference for snp association studies. *The Annals of Applied Statistics*, *4*(3), 1342. https://doi.org/10.1214/09-AOAS322
- Zellner, A., & Siow, A. (1980). Posterior oddsh ratios for selected regression hypotheses. *Trabajos de estadística y de investigación operativa*, *31*, 585–603. https://doi.org/10.1007/BF02888369

8. APPENDIX A: EXTENDED UNIT INFORMATION PRIOR

Assume a multivariate normal distribution as a prior on all edge weights, i.e., $p(\boldsymbol{\Theta} \mid \boldsymbol{\mu}, \Sigma, \gamma) \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}, \Sigma)$. Where $\boldsymbol{\mu} = (0_1, ..., 0_k)^T$ is a vector of means and Σ is a $k \times k$ covariance matrix obtained by inverting the negative of the Hessian matrix of the pseudolikelihood evaluated at the maximum pseudolikelihood.[8] Our goal is to obtain a conditional normal distribution for each edge weight $p(\theta_{ij} \mid \boldsymbol{\Theta}^{(-ij)})$ from the multivariate normal $p(\boldsymbol{\Theta} \mid \boldsymbol{\mu}, \Sigma, \gamma)$, where $\boldsymbol{\Theta}_{\boldsymbol{ij}}^{(-ij)}$ denotes the matrix of edge weights that excludes θ_{ij}. We can express this conditional distribution as $p(\theta_{ij} \mid \boldsymbol{\Theta}^{(-ij)}) \sim \mathcal{N}(\mu_{\theta_{ij}|\boldsymbol{\Theta}^{(-ij)}}, \sigma^2_{\theta_{ij}|\boldsymbol{\Theta}^{(-ij)}})$ where

$$\mu_{\theta_{ij}|\boldsymbol{\Theta}^{(-ij)}} = \mu_{\theta_{ij}} + \boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}} \times (\boldsymbol{\Sigma}_{\boldsymbol{\Theta}^{(-ij)}})^{-1} \times (\boldsymbol{\Theta}^{(-ij)} - \mu_{\theta_{ij}}),$$

note that $\mu_{\theta_{ij}} = 0$, so it can be omitted, and

$$\sigma^2_{\theta_{ij}|\boldsymbol{\Theta}^{(-ij)}} = \sigma^2_{\theta_{ij}} - \boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}} \times (\boldsymbol{\Sigma}_{\boldsymbol{\Theta}^{(-ij)}})^{-1} \times \boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}}.$$

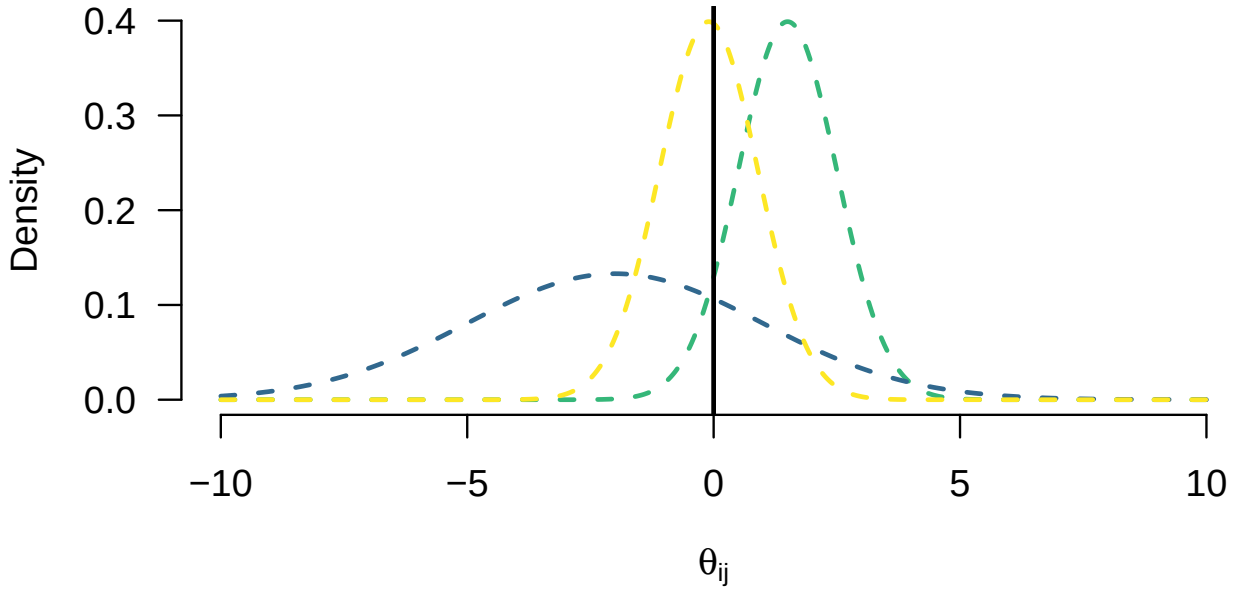
Where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{\theta_{ij}} & \boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}} \\ \boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}} & \boldsymbol{\Sigma}_{\boldsymbol{\Theta}^{(-ij)}} \end{pmatrix}.$$

Note that $\sigma^2_{\theta_{ij}}$ is a scalar, since it contains only the variance of θ_{ij}, hence we refer to it by the lowercase σ, while $\boldsymbol{\Sigma}_{\boldsymbol{\Theta}^{(-ij)}}$ is the covariance matrix for the rest of the interaction parameters excluding θ_{ij} and has the dimensions $(k-1)\times(k-1)$ and $\boldsymbol{\Sigma}_{\theta_{ij},\boldsymbol{\Theta}^{(-ij)}}$ is a vector containing the covariances between θ_{ij} and $\boldsymbol{\Theta}^{(-ij)}$. Figure 8 depicts three examples of this. Note that the slab component is now not necessarily centred at zero as it uses the conditional mean.

[8] Note that $\boldsymbol{\mu}$ should not be confused with the vector of threshold parameters.

Figure 8
A discrete spike and slab prior with three examples of an extended UI slab component.



9. APPENDIX B: SPECIFICATION OF THE LAPLACE SLAB

The MCMC sampler implemented by (MarsmanHaslbeck$_2023_ordinal$) uses a normal proposal distribution centered on the current value of θ'_{ij} and a variance based on the curvature of the pseudoposterior for the edge weight parameters. Here we explain how we obtained the variance of this pseudoposterior when implementing the Laplace distribution in a developer version of the R package bgms. The Laplace distribution is defined as follows

$$f(x) = \frac{1}{b}\exp(-\frac{|\,x-\mu\,|}{b})$$

where μ is the location parameter which is set to zero, and b is the scale parameter, whose value controls for the spread of the distribution. It is difficult to find the first and second order derivatives for the Laplace distribution due to the discontinuity at the peak of the distribution. Therefore, we implement a workaround where instead of computing the gradient and Hessian of the pseudoposterior (as is done for the UI and Cauchy priors), we directly approximate the variance of the pseudoposterior.

We obtain an approximated pseudoposterior by multiplying the normal pseudolikelihood by a prior that follows a normal distribution where its scale value is drawn from an exponential distribution with

a rate parameter of 1 i.e., $\sigma_p \sim \text{Exp}(\lambda = 1)$

$$\overbrace{\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\exp(-\frac{(x-\hat{\theta})^2}{2\hat{\sigma}^2})}^{\text{Pseudolikelihood}} \times \overbrace{\frac{1}{\sqrt{2\pi\theta_p^2}}\exp(-\frac{x^2}{2\theta_p^2})}^{\text{Normal prior}}.$$

Where $\hat{\theta}$ and $\hat{\sigma}^2$ denote the edge-specific maximum pseudolikelihood and its variance, respectively. Then, it follows that the variance of the pseudoposterior is

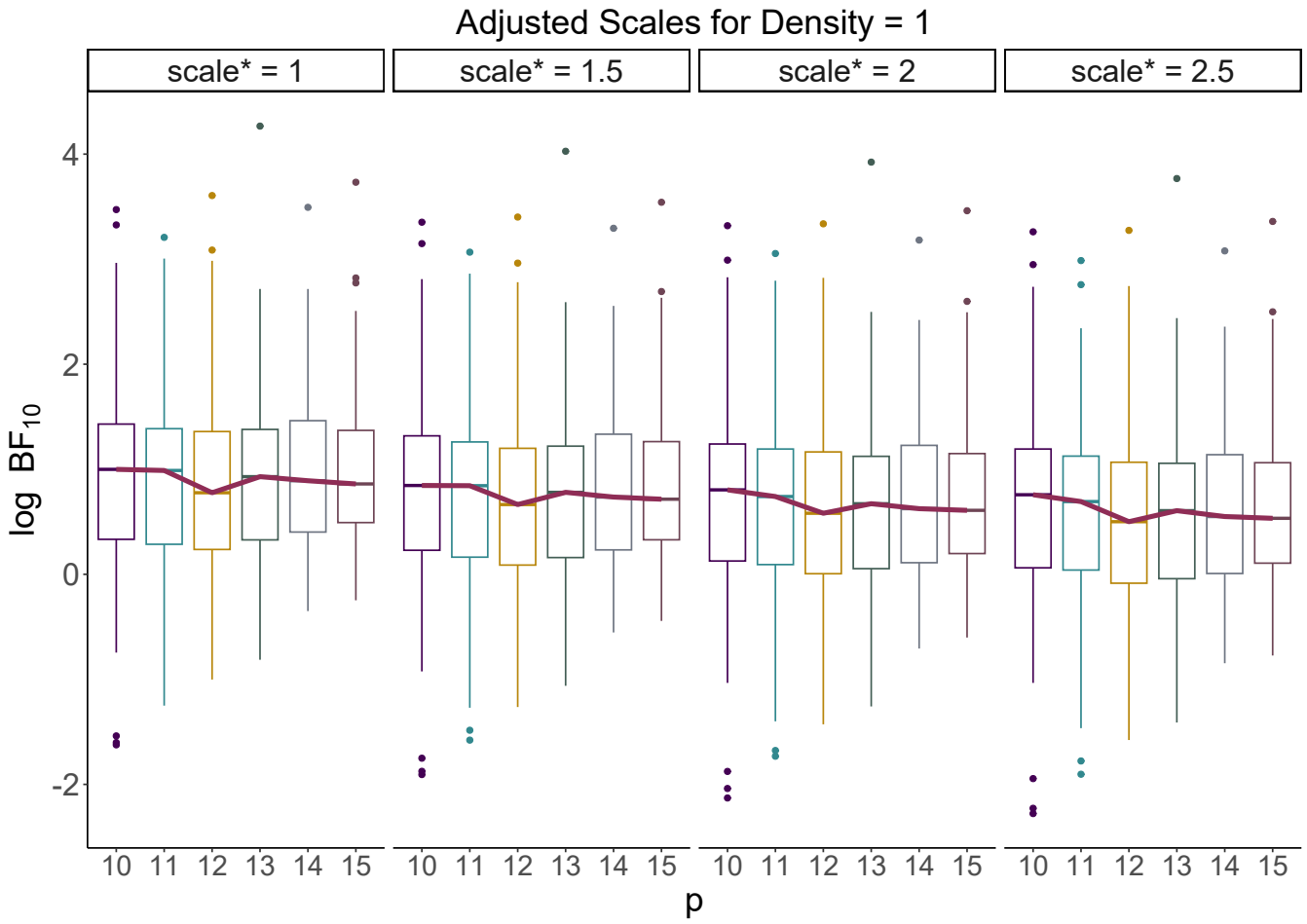
$$\theta_{pp} = \frac{\theta_p^2\hat{\theta}^2}{\theta_p^2 + \hat{\theta}^2},$$

which yields a proposal distribution of the form

$$\mathcal{N}(\theta_{ij}^{'}, \frac{\theta_p^2\hat{\theta}^2}{\theta_p^2 + \hat{\theta}^2}).$$

10. APPENDIX C: A POSSIBLE SOLUTION FOR THE SCALE OF THE CAUCHY SLAB COMPONENT WHEN THE NETWORK IS GENERATED FROM A DENSE GRAPH

Figure 9
Boxplots showing the log inclusion Bayes factors as a function of the number of variables p (the boxplots are also colored with respect to p, for easier differentiation). Different rows represent the 4 scale values for the Cauchy slab component, ordered from the lowest "scale = 1" to the highest "scale = 2.5; scale denotes the value of the old unadjusted value. The solid lines connect the median Bayes factors over p.*



We have reanalyzed the data generated from a fully dense structure by adjusting the value of the scale so that it remains constant for $p = 10$ and decreases as p increases. We achieve this by first computing a fixed transformed scale s^* as $s \times p_1^x$, where s is the original scale value (i.e., 1 to 2.5) and p_1^x is the smallest value of p (in our case 10) raised to some power x. Then, for $p_2, \ldots, p_5$, we get the new scale value s_n by ${}^{s^*}/_{p_i^x}$. For example, for $p = 10$, $p = 11$, and $p = 12$, with an original scale value of $s = 1$ and $x = 2$, the new scale values s_n would be 1, 0.83, and 0.69, respectively. For the current analysis, we varied the exponent and chose $x = 6$. Figure 9 shows that the slope of the median lines almost completely disappears compared to the unadjusted situation (cf. Figure 7).[9] These results show that the scale of

[9] Note that the slight dip at $p = 12$ is due to some fluctuations in the data generation process and is not related to the scaling.

the Cauchy slab component must be reduced in a nonlinear manner to obtain comparable inclusion Bayes factors. However, this approach can only be used when the graph structure is (fully) dense.