MULTIVARIATE BEHAVIORAL RESEARCH
2024, VOL. 59, NO. 5, 913–933

Testing Conditional Independence in Psychometric Networks: An Analysis of Three Bayesian Methods

Nikola Sekulovski[a], Sara Keetelaar[a], Karoline Huth[a,b,c], Eric-Jan Wagenmakers[a], Riet van Bork[a], Don van den Bergh[a,d], and Maarten Marsman[a,c,d]

[a]Department of Psychology, University of Amsterdam, Netherlands; [b]Department of Psychiatry, Amsterdam UMC Location, University of Amsterdam, Netherlands; [c]Centre for Urban Mental Health, University of Amsterdam, Netherlands; [d]Amsterdam Brain and Cognition, University of Amsterdam, Netherlands

ABSTRACT

Network psychometrics uses graphical models to assess the network structure of psychological variables. An important task in their analysis is determining which variables are unrelated in the network, i.e., are independent given the rest of the network variables. This conditional independence structure is a gateway to understanding the causal structure underlying psychological processes. Thus, it is crucial to have an appropriate method for evaluating conditional independence and dependence hypotheses. Bayesian approaches to testing such hypotheses allow researchers to differentiate between absence of evidence and evidence of absence of connections (edges) between pairs of variables in a network. Three Bayesian approaches to assessing conditional independence have been proposed in the network psychometrics literature. We believe that their theoretical foundations are not widely known, and therefore we provide a conceptual review of the proposed methods and highlight their strengths and limitations through a simulation study. We also illustrate the methods using an empirical example with data on Dark Triad Personality. Finally, we provide recommendations on how to choose the optimal method and discuss the current gaps in the literature on this important topic.

KEYWORDS

Bayesian model averaging; Bayes factor; Conditional independence; Markov random fields

Introduction

In network psychometrics, graphical models known as Markov Random Fields (Kindermann & Snell, 1980; Rozanov, 1982) have become a popular tool for assessing the network structure of psychological variables (see Borsboom et al., 2021; Contreras et al., 2019; Marsman & Rhemtulla, 2022; Robinaugh et al., 2020, for recent reviews). In these networks, nodes represent observed psychological variables (e.g., symptom indicators), and edges represent the pairwise relationships between them. A 'present' relationship indicates a direct link (positive or negative) between a pair of variables, while an absent relationship reflects conditional independence; if any dependence exists it is fully accounted for by other variables in the network. For example, although shark attacks (A) and ice cream sales (I) are correlated, their dependence disappears when season (S) is taken into account, showing that A and I are conditionally independent given S. The goal of network analysis is to discover the underlying configuration of present and absent edges—the conditional independence structure of the network. Conditional independence is a first step in determining causal relationships (e.g., Pearl, 2009; Spirtes et al., 2000), and aids our understanding of the underlying dynamic system. With the gradual development of Bayesian approaches to network analysis (e.g., Marsman et al., 2015; Marsman & Haslbeck, 2023; Mohammadi & Wit, 2015; Williams, 2021; Williams & Mulder, 2020a), testing the conditional independence of a pair of variables in the network has also begun to receive attention: Three Bayesian approaches to conditional independence testing have recently been proposed in the network psychometrics literature. Although the development of conditional independence tests is an important step forward, the three methods differ conceptually, and we believe that their foundations are not well known. In this paper, we provide a conceptual review of the three Bayesian approaches, and show

that when we are uncertain about the structure of the network, there appears to be an optimal approach for testing conditional independence.

In recent years, both frequentist and Bayesian approaches have been proposed for estimating the network structure from empirical data. In psychology, frequentist approaches are the norm, and although unregularized estimation approaches have become available in recent years (see Blanken et al., 2022, for a review), regularized estimation based on lasso remains popular (e.g., Borsboom et al., 2021; Epskamp & Fried, 2018; Tibshirani, 1996; van Borkulo et al., 2014). A caveat of these approaches is that they are biased toward the null hypothesis of conditional independence, yet their underlying frequentist methodology can only refute this hypothesis, not support it. This leads to the problem that if an edge is missing from an estimated network, we are unsure whether this is due to a lack of information in the data to support the relation, or due to actual conditional independence (e.g., Epskamp et al., 2017). Bayesian methods (e.g., Wagenmakers, Marsman, et al., 2018; Wagenmakers et al., 2016), on the other hand, are able to quantify the relative support for the competing hypotheses of conditional dependence and conditional independence; importantly, they can also reveal the lack of support for either hypothesis in the data at hand. By distinguishing evidence of absence from absence of evidence, Bayesian methods facilitate a deeper understanding of the conditional independence structure of the network.

Below we review three Bayesian approaches that have recently been proposed to test conditional independence. The first method uses the credible interval—the Bayesian version of the frequentist confidence interval—and assesses whether or not it contains the parameter values that indicate conditional independence. This method focuses solely on rejecting the conditional independence hypothesis, and thus suffers from the same fundamental problem that plagues the frequentist methods mentioned above. The second method uses a Bayes factor approach (Jeffreys, 1961; Kass & Raftery, 1995)—which is the Bayesian generalization of the likelihood ratio test. The Bayes factor compares how well two competing models can predict the observed data. When we compare two models that are identical except that one has two variables that are unrelated and the other has them related, we can use the Bayes factor to express the relative support, or lack thereof, for the conditional dependence or independence hypotheses. The Bayes factor test represents a major improvement over interval-based tests for conditional dependence and independence. However, we will show that this Bayes factor approach requires a choice about which relationships are present in the rest of the network and that it is sensitive to that choice. The third method, called the inclusion Bayes factor, is a generalization of the Bayes factor approach that uses Bayesian model averaging (BMA, Hoeting et al., 1999; Kaplan, 2021) to overcome the sensitivity to which relationships are present in the rest of the network. The inclusion Bayes factor compares how well we can predict the observed data from a combination of all models in which the two variables are related, and compares this to the predictive adequacy of a combination of models in which the variables are unrelated. In this paper, we consider the structure of the network, a particular configuration of present and absent edges, to be a model. As we will show below, BMA allows us to make robust structure-averaged inferences.

The remainder of this paper is structured as follows. The next two sections provide a conceptual introduction to the role of conditional dependence and independence in graphical modeling, and the Bayesian methodology that underlies the methods for testing these hypotheses. We refer the interested reader to Epskamp et al. (2022), Marsman et al. (2018), and Waldorp and Marsman (2022) for a detailed introduction to the graphical models used in network psychometrics, to van de Schoot et al. (2014) and Wagenmakers, Marsman, et al. (2018) for a detailed introduction to Bayesian estimation and hypothesis testing, and to Huth, de Ron, et al. (2023) for a more comprehensive introduction to the Bayesian analysis of graphical models. In the third section, we investigate the three Bayesian methods to test for conditional dependence and independence in detail and discuss their relations and limitations, after which we compare their relative performance in a simulation study and illustrate them with an empirical example. We end by discussing the limitations of the Bayesian analysis of graphical models, and Bayesian model averaging in particular.

Graphical modeling

A graphical model specifies the joint probability distribution for a set of observed variables, and represents these variables as nodes in a network. The goal of a statistical analysis of the graphical model is to determine the relations between pairs of variables, which will constitute the edges of the network. We usually have two questions about network relations. First, we wish to know if the edge is there or not: *is*

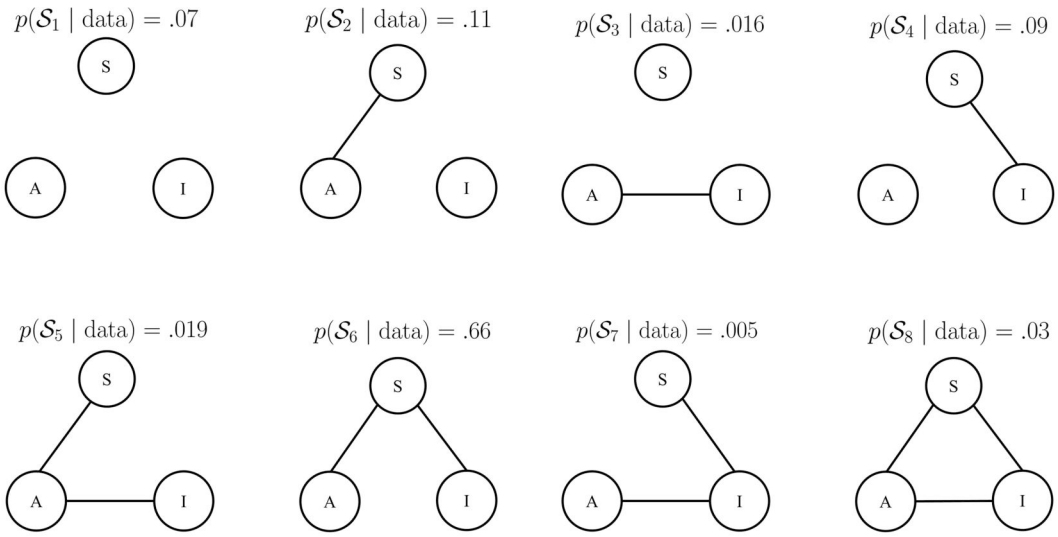

Figure 1. The possible structures along with their Posterior Structure Probabilities for the random variables: shark attacks (A), ice cream sales (I), and season (S).

an effect present? Once we have established that there is an effect, and the edge should be in the model, a follow-up question could be how strong the relation is: *what is the strength of the effect?* The first question is usually linked to testing whereas the latter is linked to estimation, but this distinction can be vague in practice.

In this paper, we will address the two questions about the graphical model separately. First, a binary variable γ_{ij} is used to indicate that the edge between variables i and j is present (i.e., $\gamma_{ij} = 1$) or absent (i.e., $\gamma_{ij} = 0$). In a network with p variables, we have $k = p\ (p-1)/2$ possible edges. Each configuration of edges (i.e., the pattern of zeros and ones: γ_{12}, ..., $\gamma_{(p-1)p}$) constitutes a possible network structure $\mathcal{S}_s$, of which there are 2^k in total. Figure 1 illustrates the idea for a network of three random variables; shark attacks (A), ice cream sales (I), and season (S). The three variables yield $k = 3 \times (3-1)/2 = 3$ possible edges, and thus $2^k = 2^3 = 8$ possible network structures. For example, Structure $\mathcal{S}_1 = [\gamma_{AI} = 0,\ \gamma_{AS} = 0,\ \gamma_{SI} = 0]$ and Structure $\mathcal{S}_4 = [\gamma_{AI} = 0,\ \gamma_{AS} = 0,\ \gamma_{SI} = 1]$.

For each structure $\mathcal{S}_s$, $s = 1$, ..., 2^k, we have a distinct statistical model[1] for the observed variables $p(\text{data}|\boldsymbol{\Theta}_s,\ \mathcal{S}_s)$. The weights of the relations in the structure $\mathcal{S}_s$ are expressed in a symmetric matrix $\boldsymbol{\Theta}_s$, which has the edge weights θ_{ij} on the off-diagonals. The edge weights that correspond to relations that are absent from $\mathcal{S}_s$ are set to zero in $\boldsymbol{\Theta}_s$. For the graphical models that we analyze in this paper—Markov Random Field (MRF) models (Kindermann & Snell, 1980; Rozanov, 1982)—these edge weights are partial associations, which indicate the strength of the relation between two variables that excludes the influence of other variables in the model. The higher the absolute value of the partial association θ_{ij}, the stronger the two variables influence each other.

Several MRF models are used in network psychometrics, which differ primarily in the level of measurement of the variables. For example, the Ising model (Ising, 1925) is a graphical model for binary variables (e.g., symptom indicators), the ordinal MRF (Marsman & Haslbeck, 2023) extends the Ising model to also include ordinal variables, the Gaussian graphical model (GGM; Lauritzen, 2004) is used for continuous variables, and the mixed graphical model (MGM; Haslbeck & Waldorp, 2020) handles binary, unordered categorical, count, and continuous variables. For the particular case of the GGM, the matrix $\boldsymbol{\Theta}$ is known as the precision matrix, and when standardized, it contains partial correlations (e.g., Waldorp & Marsman, 2022). To keep the discussion general, we will refer to the elements of $\boldsymbol{\Theta}$ as edge weights. Other graphical models that are also used in the network psychometrics literature but are not MRF models are the multivariate ordered probit (Guo et al., 2015) for ordinal variables and the Gaussian copula graphical model (Dobra & Lenkoski, 2011) for mixed binary, ordinal, and continuous variables. However, in this paper we focus exclusively on MRF models.

Conditional independence

In network psychometrics, it is often assumed that the observed data are variables in a complex, dynamic

[1]In the Bayesian literature, the model is usually indicated with $\mathcal{M}_s$, but here we use $\mathcal{S}_s$ to connect it to the network's structure.

system. The underlying system has a causal component in that some variables influence other variables in a particular way, and some of these relationships are reciprocal. Since it is difficult to learn the directed, causal relationships from correlational data, we use undirected graphical models to model the relationships among the variables in the underlying system. MRFs are an important class of undirected graphical models because their parameters tell us directly about the conditional dependence and independence between variables in the network: If the edge weight θ_{ij} between variables i and j is zero, then the two variables are conditionally independent. MRFs are thus convenient models for assessing conditional independence, and, since conditional independence is a gateway to learning the underlying causal structure (e.g., Pearl, 2009; Spirtes et al., 2000), they play an important role in the graphical approach to causal inference (Ryan et al., 2022). One could, of course, adopt a purely statistical interpretation of conditional independence without considering potential causal implications. However, since the notion of conditional independence is also central to causal inference, we wish to clarify how the two are related in this subsection.

Spirtes et al. (2000), Pearl (2009), and others (see Glymour et al., 2019, for a recent overview) have developed the graphical approach to causal inference as a formal framework in which causal relationships are represented as directed acyclic graphs (DAGs)0.2 Conditional dependencies and independencies are key to identifying DAGs that are consistent with observed data. For example, consider the three variables A, S, and I in Figure 1. From their correlations alone, we cannot identify causal relationships among the three variables. However, if we also knew their conditional dependencies, e.g., that A and I are conditionally independent given S, while A and S, and A and I are conditionally dependent (i.e., A–S–I, such as $\mathcal{S}_6$ in Figure 1), we could take a step toward causal discovery. Under some strong assumptions (e.g., there are no unobserved confounders, there is no selection bias, and the causal relations do not cancel each other out; Eberhardt, 2017), one can use the conditional independence structure $\mathcal{S}$ to infer three possible (directed) causal graphs: A → S → I, A ← S ← I, and A ← S → I. For a detailed introduction to learning causal relations from conditional dependence and independence, we refer the interested reader to Pearl (2009).

The conditional independence structure $\mathcal{S}$ is a middle ground between simple unconditional associations and directed causal graphs: Simple associations will contain many spurious relations that disappear when conditioning on other variables in the network. While this conditioning removes associations that can be explained through other variables in the network, it can also induce spurious relations: any variable that is a common effect of other variables in the network will induce a spurious association between these variables when conditioned on. It is therefore important to note that not all conditional dependencies will reflect causal relations unless strong assumptions are made (such as the absence of common effects and unobserved common causes). But, the conditional dependency structure will contain conditional dependencies for every causal relation in the causal graph. In this sense, the conditional independence structure can generate possible *hypotheses* about causal paths, but cannot be used to infer causal paths directly (see Ryan et al., 2022, for a more detailed discussion of the problems of causal inference from network models). But, for those who do want to take a next step and identify directed causal graphs, causal discovery is an exciting field with many advances, such as causal discovery algorithms that do not require the absence of unobserved common causes or feedback loops (Eberhardt, 2017).

There are at least three reasons for why one might want to model the conditional independence structure of the MRF rather than going a step further and using the MRF to discover directed causal graphs. First, inferring a DAG[2] from conditional dependencies in observational data requires strong assumptions that may not hold in practice (e.g., no unobserved common causes and no feedback loops). Second, for a conditional independence structure, there may be many directed causal graphs that are equivalent and consistent with the conditional independence structure. We have already seen that there are several equivalent graphs for the three-variable example above, and for more than three variables the set of equivalent graphs increases enormously. Therefore, it may be much easier to work with a single MRF than with the potentially large set of equivalent causal graphs (Epskamp et al., 2018). Third, the MRF does not commit one to a causal interpretation; instead, one can choose a purely statistical interpretation of predicting variables from other variables in the network or other interpretations (e.g., Epskamp et al., 2022).

[2]DAGs are sometimes referred to as Bayesian networks. We wish to emphasize that Bayesian networks (DAGs) are different from Bayesian analysis of (MRF) graphical models, which is the focus of this paper.

Bayesian graphical modeling

Bayesian inference aims to use data to update our knowledge about the network structure $\mathcal{S}_s$—the collection of edges in the network—and the network parameters $\boldsymbol{\Theta}_s$—the edge weights. To allow the data to update our knowledge of the network structure and parameters, we need to make explicit what we know about them before seeing the data. Figure 1 shows that there are many possible structures that that could underlie the network, and similarly, there are many possible values for the corresponding edge weights. But which values could describe our data? Since our goal is to learn about them, the specific configuration of the network relations and the exact parameter values are usually unknown to us. To account for this uncertainty, we assign prior distributions to the model or structure $\mathcal{S}_s$ and to the parameters of that model $\boldsymbol{\Theta}_s$. A prior is a probability distribution that a Bayesian uses to assign weights (i.e., probability or probability density) to different values of the parameters and structure. First, we assign prior probabilities to the different network structures $p(\mathcal{S}_s)$ (i.e., the prior distribution of the effect), and then, conditional on a particular structure, we specify prior distributions on the corresponding edge weights $p(\boldsymbol{\Theta}_s \mid \mathcal{S}_s)$ (i.e., the prior distribution of the effect size). The priors provide a way to formalize theory and incorporate advanced knowledge (e.g., results from previous research; Lindley, 2004; Vanpaemel & Lee, 2012), or they can be used to express ignorance using a default or objective prior specification (e.g., Consonni et al., 2018). In the appendix, we provide details about the prior distributions implemented in three popular R packages for analyzing MRF models.

Regardless of how we specify the priors, Bayes' rule weighs the prior distribution with the information coming from the observed data to update it to a posterior distribution,

$$p(\boldsymbol{\Theta}_s,\ \mathcal{S}_s \mid \text{data}) = \frac{p(\text{data} \mid \boldsymbol{\Theta}_s,\ \mathcal{S}_s)p(\boldsymbol{\Theta}_s \mid \mathcal{S}_s)p(\mathcal{S}_s)}{p(\text{data})}.$$

This joint posterior distribution expresses everything that we know about the structure *and* parameter values of the network after seeing the data and is central to the Bayesian analysis of graphical models. The different Bayesian tests for conditional independence consider different aspects of this joint posterior. To make this more explicit, we factor the joint posterior as follows

$$p(\boldsymbol{\Theta}_s,\ \mathcal{S}_s \mid \text{data}) = p(\boldsymbol{\Theta}_s \mid \mathcal{S}_s,\ \text{data}) \times p(\mathcal{S}_s \mid \text{data}),$$

and express it as a product of the posterior distribution of the parameters $\boldsymbol{\Theta}_s$ under the specific structure $\mathcal{S}_s$, and the posterior distribution of the possible structures with the parameters integrated out. The former is referred to as the conditional posterior distribution for the network parameters (i.e., it is the posterior distribution of the edge weights for a specific structure $\mathcal{S}_s$) and the latter as the marginal posterior distribution of the network structure (i.e., a posterior of the structures without the parameter values for the edge weights). Below, we will use the conditional posterior distribution $p(\boldsymbol{\Theta}_s \mid \mathcal{S}_s, \text{data})$ for Bayesian parameter estimation, and the marginal posterior distribution $p(\mathcal{S}_s \mid \text{data})$ for Bayesian hypothesis testing.

Bayesian hypothesis testing: the Bayes factor

Two out of the three proposed Bayesian methods for testing the conditional independence hypothesis that we review in the next section make use of the Bayes factor (Jeffreys, 1939; Kass & Raftery, 1995). The Bayes factor quantifies the relative predictive performance of two rival hypotheses (e.g., the conditional dependence of two variables or their conditional independence), or of two competing models or structures. Consider two competing network structures $\mathcal{S}_s$ and $\mathcal{S}_t$. The Bayes factor is defined as the change in beliefs concerning the relative plausibility of the two structures before and after observing the data

$$\underbrace{\frac{p(\mathcal{S}_s)}{p(\mathcal{S}_t)}}_{\substack{\text{Prior}\\ \text{odds}}} \times \underbrace{\frac{p(\text{data}|\mathcal{S}_s)}{p(\text{data}|\mathcal{S}_t)}}_{\text{BF}_{st}} = \underbrace{\frac{p(\mathcal{S}_s|\text{data})}{p(\mathcal{S}_t|\text{data})}}_{\substack{\text{Posterior}\\ \text{odds}}}.$$

Specifically, the first factor on the left of the formula above is the prior odds, that is, the relative plausibility of the two structures before having seen the data. The second factor is the Bayes factor which indicates the statistical evidence or support for the two structures in the data at hand. The term on the right is the posterior odds, which indicates the relative plausibility of the rival models after having seen the data. In this paper, we assume that the prior odds are equal to one by assuming $p(\mathcal{S}_s) = p(\mathcal{S}_t)$, which makes the Bayes factor equal to the posterior odds (see Marsman et al., 2022, for a different approach).

The subscripts in the Bayes factor notation indicate in which direction the support is expressed. BF_{st} indicates the relative support for $\mathcal{S}_s$ over $\mathcal{S}_t$ and BF_{ts} indicates the relative support for $\mathcal{S}_t$ over $\mathcal{S}_s$. Observe that the Bayes factor BF_{ts} is the reciprocal of BF_{st}, i.e., $\text{BF}_{ts} = 1/\text{BF}_{st}$. The Bayes factor BF_{st} ranges from 0 to ∞, values larger than one indicate a relative support for $\mathcal{S}_s$ while values smaller than one indicate the

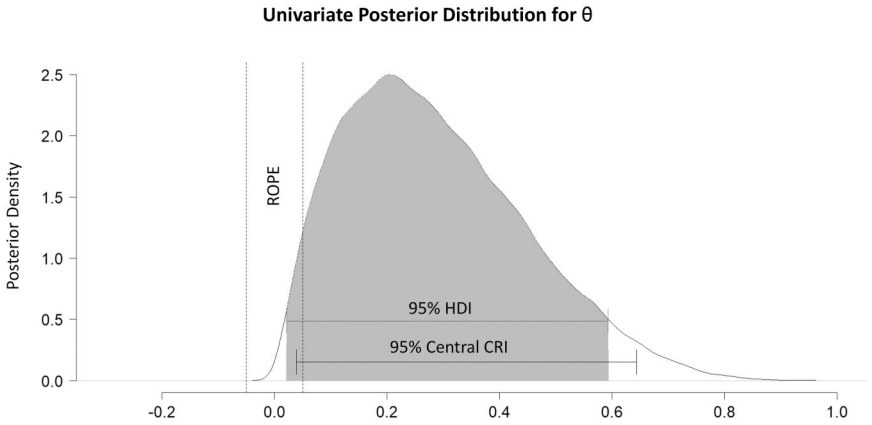

Figure 2. An example of a posterior distribution for a parameter θ, the line at the bottom of the density represents the 95% central credible interval and the shaded gray region represents the 95% highest density interval (HDI). The two dashed vertical lines around zero represent the region of practical equivalence (ROPE, introduced in the next section).

relative support for $\mathcal{S}_t$. If the Bayes factor is equal to one, both structures predicted the data equally well. In practice, we usually interpret Bayes factors between 1/10 and 10 as evidence that is insufficiently compelling.[3]

Bayesian estimation: the posterior distribution of partial associations

One of the three Bayesian methods for testing the conditional independence hypothesis that we review in the next section makes use of Bayesian estimation. In practical situations, we often wish to estimate the parameters $\boldsymbol{\Theta}'$ for a particular structure $\mathcal{S}'$. This could be the structure with the highest posterior probability $p(\mathcal{S}' \mid \text{data})$, the median probability structure (e.g., Barbieri & Berger, 2004; Marsman et al., 2022) that consists of all relations for which the posterior inclusion probability (defined in the section on Bayesian model averaging) is greater than a half, or it could be the complete structure that includes all relations. The posterior distribution for a single structure $\mathcal{S}'$ is

$$p(\boldsymbol{\Theta}' \mid \mathcal{S}',\ \text{data}) = \frac{p(\text{data} \mid \boldsymbol{\Theta}',\ \mathcal{S}') \times p(\boldsymbol{\Theta}' \mid \mathcal{S}')}{p(\text{data} \mid \mathcal{S}')},$$

where $p(\boldsymbol{\Theta}' \mid \mathcal{S}')$ denotes the prior distribution for the parameters under the structure $\mathcal{S}'$. Given a single parameter θ_{ij}, the prior $p(\theta'_{ij} \mid \mathcal{S}')$ assigns a relative plausibility to each value of the parameter. The information in the data is then used to update this prior distribution into a posterior distribution $p(\theta_{ij} \mid \mathcal{S}', \text{data})$. In the posterior distribution, the plausibility of parameter values that predict the data well increases, while the plausibility of parameter values that predict the data poorly decreases (Wagenmakers et al., 2016).

Instead of reporting the full posterior distribution for each element in $\boldsymbol{\Theta}'$, we often report it in terms of a measure of location (i.e., the posterior mean, median, or mode) and spread (i.e., the posterior variance), or in terms of an $x\%$ credible interval (see van Doorn et al., 2021). An $x\%$ credible interval contains $x\%$ of the probability mass of the posterior distribution. Two popular ways to create an $x\%$ credible interval are the highest posterior density interval, which is the shortest possible credible interval that contains $x\%$ of the posterior mass, and the $x\%$ central credible interval, which is obtained by clipping $(100 - x)/2\%$ from each tail of the posterior distribution. Figure 2 shows a fictional example of a posterior distribution that has a 95% central credible interval and a 95% highest density interval. The posterior is a probability density with the gray area under its curve containing 95% of its total probability. Note that the highest density interval is shorter than the central credible interval, even though both capture 95% of the posterior.

Equipped with these Bayesian concepts, we next turn to the three proposed Bayesian approaches for testing conditional independence.

Three Bayesian methods for testing conditional independence

Approach 1: Credible interval

In frequentist statistics, an assessment of whether or not the null value θ_0 falls within the $x\%$ confidence interval for a parameter θ_{ij} (sometimes considered as estimation, but see Morey et al., 2016) is equivalent to the test of the null hypothesis,

[3]In principle, Bayes factors are a continuous measure of evidence and therefore do not require strict cutoff values. But even if we do, there is no hard and fast rule for what the cutoff should be, and practitioners may prefer other values (Jeffreys, 1961; Kass & Raftery, 1995).

$$\mathcal{H}_0 : \theta_{ij} = \theta_0,$$

with a significance level of $\alpha = (100 - x)\%$: We would reject $\mathcal{H}_0$ with significance level α if the null value falls outside the $(100 - \alpha)\%$ confidence interval (cf. Figure 2). It is tempting to extend this testing approach to Bayesian statistics by using an $x\%$ credible interval to test whether or not we could reject $\mathcal{H}_0$. But from which posterior distribution should we take the credible interval for the partial associations? In practice, this is usually done using a complete structure $\mathcal{S}_C$ that includes all relations (e.g., the bottom right structure in Figure 1). However, this approach implies that the relation between nodes i and j is *a priori* assumed to exist, and we are thus testing a hypothesis that we assume to be false from the outset (e.g., Jeffreys, 1939). This signals a bias against the null hypothesis, which is common in classical null hypothesis significance tests.

In practice, the logic behind credible interval-based tests may indeed lead to contradictions, since comparisons between the null hypothesis and its complement using the Bayes factor, for example, may signal support for the null hypothesis of conditional independence, while the null value θ_0 would fall outside the credible interval. See Berger and Delampady (1987) and Wagenmakers et al. (2020) for detailed discussions of this issue. Null hypothesis tests based on the credible interval can also lead to ambiguous results, because if the null value would fall within the interval, we cannot interpret this as support for the null hypothesis because the test cannot distinguish between the potential causes of this failure to reject (i.e., absence of evidence or evidence of absence). In order to test for conditional independence, we must therefore be able to quantify support in favor of the null hypothesis.

Despite this complication, credible interval-based tests have been used to test for conditional independence in the Bayesian graphical modeling literature. For example, Jongerling et al. (2023) use credible intervals to perform edge selection (i.e., conditional independence testing) in GGMs with the goal of estimating the posterior distribution of centrality measures. Williams (2021) used a generalization of the credible interval test based on the idea that we can specify a region in the parameter space that is essentially zero—the region of practical equivalence (ROPE, Kruschke, 2011)—and then exclude an edge if $x\%$ of the posterior distribution of the partial association is inside the ROPE, otherwise include it (cf. Figure 2). In a slightly different way, Marsman et al. (2022) also used credible intervals for edge selection. They used a continuous spike and slab prior on the partial associations of an Ising model, where the intersection of the spike and slab components occurs at an approximate $x\%$ credible interval. This is very similar to using ROPE for edge selection; to set the spike-and-slab prior, Marsman et al. (2022) also start with a posterior distribution that assumes the effect is present (based on the unit-information prior; Kass & Wasserman, 1995). However, unlike the credible interval test and the ROPE approach, the approach in Marsman et al. (2022) can distinguish the potential causes underlying the edge exclusion because it assigns prior weights to the edge inclusion and exclusion hypotheses.

Thus, our concerns with credible interval-based tests are directed at their conceptual underpinnings, particularly their inability to quantify support for the null hypothesis. To quantify this support, we need an evidential measure that contrasts the competing hypotheses of conditional dependence and independence, i.e., the Bayes factor. For a review of (log) Bayes factors as weight of evidence see, for example, Good (1985).

Approach 2: The single-model Bayes factor

The Bayes factor is the gold standard for Bayesian hypothesis testing (Berger & Pericchi, 2015), and around the same time that graphical models became popular in psychology, Bayes factor hypothesis testing became popular in psychological research. In large part, this increased popularity of the Bayes factor in psychology is a response to the misuse of the null hypothesis significance test (NHST) in psychological research and the limited replicability (Ioannidis, 2005; Open Science Foundation, 2015) of many psychological findings established with NHST (e.g., Wagenmakers, 2007; Wagenmakers et al., 2011). Some of the concerns that methodologists have with NHST also play a role in the credible interval test of the previous section. One of the more prominent concerns is that the adequacy or inadequacy of the null hypothesis is not compared against an alternative. Thus, rejection of the null hypothesis should not be taken as evidence in favor of the alternative hypothesis, which may be just as inadequate as the null hypothesis (or even more inadequate). The Bayes factor, however, compares the predictive adequacy of the null hypothesis against that of an alternative, and as such can separate evidence for the absence of an effect, evidence for the presence of an effect, but also the absence of evidence in either direction (e.g., Dienes, 2014; Keysers et al., 2020). Thus, Bayes factor testing is a significant step forward for

psychological network analysis. However, we are concerned with the way it is commonly formulated.

We consider the Bayes factor test for the conditional independence of the variables i and j in the network, i.e., we consider the following two hypotheses

$$\mathcal{H}_0 : \theta_{ij} = 0, \text{ and } \mathcal{H}_1 : \theta_{ij} \neq 0.$$

If we wish to assign prior probabilities to these hypotheses, it is easier to reformulate them in terms of the edge indicators and model the size of the effect θ_{ij} conditional on the presence of the effect. That is, we model $p(\boldsymbol{\Theta}_s|\mathcal{S}_s)$ such that we can impose a prior on the hypothesis or model $\mathcal{S}_s$. Then, our hypotheses can be reformulated as

$$\mathcal{H}_0 : \gamma_{ij} = 0, \text{ and } \mathcal{H}_1 : \gamma_{ij} = 1.$$

By formulating the hypothesis in terms of the edge indicator rather than the edge weight, we immediately encounter a problem. We cannot yet isolate the effect of a single relationship, i.e., the edge indicator, and thus we must now carefully consider how to set up the Bayes factor. The way this is usually done is by comparing two structures $\mathcal{S}_s$ and $\mathcal{S}_t$ that are identical except that the relation between the variables i and j is present in $\mathcal{S}_s$ but is absent in $\mathcal{S}_t$. In this way, comparing $\mathcal{S}_s$ with $\mathcal{S}_t$ using the Bayes factor gives us a Bayes factor test for $\mathcal{H}_1$ versus $\mathcal{H}_0$. Although it is not made explicit, in practice the complete structure $\mathcal{S}_s = \mathcal{S}_C$ is used here, as in the case of the credible interval test we discussed earlier.

Note that the Bayes factor test for conditional independence we formulated above is not uniquely defined. In principle, we could compare any two structures $\mathcal{S}_s$ and $\mathcal{S}_t$, as long as they are identical except that the relation between the variables i and j is present in $\mathcal{S}_s$ but not in $\mathcal{S}_t$. For our hypothetical three-variable example, this means that we have three ways to test the conditional independence of variables A and I: We could contrast $\mathcal{S}_s = \mathcal{S}_3$ with $\mathcal{S}_t = \mathcal{S}_1$, $\mathcal{S}_s = \mathcal{S}_5$ versus $\mathcal{S}_t = \mathcal{S}_2$, or $\mathcal{S}_s = \mathcal{S}_8$ versus $\mathcal{S}_t = \mathcal{S}_6$. Each of these comparisons is a valid comparison in terms of contrasting the effect of the relation (i.e., assessing conditional independence). However, in each case we are making a different assumption about the other relationships in the network. We will refer to any such Bayes factor test as a single-model Bayes factor, since it assumes a single model for the remaining relationships in the network. The single-model Bayes factor test is sensitive to the assumption concerning the overall network structure because partial associations are sensitive to the other partial associations in the model or structure. To illustrate, consider the relation between variables A and I in our three-variable example. First, in order to compute its value in our example, we express the Bayes factor as a function of the prior and posterior probabilities:

$$\text{BF}_{st} = \underbrace{\frac{p(\mathcal{S}_s \mid \text{data})}{p(\mathcal{S}_t \mid \text{data})}}_{\substack{\text{Posterior}\\ \text{odds}}} \Big/ \underbrace{\frac{p(\mathcal{S}_s)}{p(\mathcal{S}_t)}}_{\substack{\text{Prior}\\ \text{odds}}}.$$

The posterior probabilities for each of the eight structures are shown in Figure 1. When we assume that each of the structures is equally plausible a priori, the prior probabilities are equal to one and the Bayes factors are equal to the posterior probabilities. Thus, the Bayes factors for the three possible model pairs are obtained as follows:

$$\begin{aligned}
\text{BF}_{31} &= \frac{p(\mathcal{S}_3|\text{data})}{p(\mathcal{S}_1|\text{data})} = \frac{.016}{.07} = 0.23,\\
\text{BF}_{52} &= \frac{p(\mathcal{S}_5|\text{data})}{p(\mathcal{S}_2|\text{data})} = \frac{.019}{.11} = 0.17,\\
\text{BF}_{86} &= \frac{p(\mathcal{S}_8|\text{data})}{p(\mathcal{S}_6|\text{data})} = \frac{.03}{.66} = 0.05.
\end{aligned}$$

This demonstration confirms that the single-model Bayes factor can, in fact, be sensitive to our choice for the remaining relations in the network. The first two Bayes factors (i.e., $\text{BF}_{31} = 0.23$ and $\text{BF}_{52} = 0.17$) showed weak evidence for exclusion, while the third Bayes factor showed strong evidence for exclusion (i.e., $\text{BF}_{86} = 0.05$). But which Bayes factor test should we use?

Williams and Mulder (2020a) proposed the single-model Bayes factor for testing conditional independence in GGMs (see also Giudici, 1995). In their approach, the complete structure is used as a basis for comparison. In the next section, we show that this method works well when the data generating structure has relatively many relations, consistent with the model's assumption, but it starts to perform less well when the data generating structure is sparse and has relatively few connections. Since we are typically highly uncertain about which particular structure would underlie our data (see Marsman et al., 2022; Marsman & Haslbeck, 2023), the foundations of the single-model Bayes factor can be unstable.

Approach 3: The inclusion Bayes factor

We can use Bayesian model averaging (BMA; Hoeting et al., 1999; Kaplan, 2021) to overcome the sensitivity of the single-model Bayes factor to our assumptions about the remaining relationships in the network. When we consider the single-model Bayes factor, we

must assume that the network is based on some structure. In practice, however, we usually do not know what that structure is. To account for our uncertainty about the structure of the network, BMA considers all possible structures and weights the outcome of each structure by its posterior probability; the relative plausibility that the structure produced the data at hand. By weighting the outcome of each structure by its posterior probability, BMA accounts for the uncertainty we have about which structure is at play (Hinne et al., 2020; Huth, de Ron, et al., 2023). Mohammadi and Wit (2015) and Marsman et al. (2022) applied BMA to graphical models.

We focus here on the posterior inclusion probability, the posterior probability of including an effect, which we use to estimate the inclusion Bayes factor; the Bayes factor test that pits the conditional dependence hypothesis against the conditional independence hypothesis. Although we do not consider it here, BMA is also useful for estimating the marginal posterior distribution for the partial associations; a robust estimate of the effect size that incorporates the uncertainty in the parameter and the uncertainty in its selection.

We can express the posterior probability of including the edge between variables i and j as the sum of the posterior probabilities over all structures that include the edge. Let $\mathcal{S}^{(i-j)}$ denote the set of structures that include an edge between variables i and j, then the inclusion probability can be computed as

$$p(\gamma_{ij} = 1|\text{data}) = \sum_{\mathcal{S}' \in \mathcal{S}^{(i-j)}} p(\mathcal{S}'|\text{data}),$$

which weights the posterior plausibility of the inclusion of the relation in the network structure. For example, the posterior inclusion probability of including the relation between variables A and I (i.e., $\gamma_{AI} = 1$) in Figure 1 is equal to

$$\begin{aligned} p(\gamma_{AI} = 1|\text{data}) &= p(\mathcal{S}_3|\text{data}) + p(\mathcal{S}_5|\text{data}) \\ &\quad + p(\mathcal{S}_7|\text{data}) + p(\mathcal{S}_8|\text{data}) \\ &= .016 + .019 + .005 + .03 = .07. \end{aligned}$$

Since the posterior probabilities for edge inclusion and exclusion sum to one, the corresponding probability of exclusion is $p(\gamma_{AI} = 0|\text{data}) = 1 - p(\gamma_{AI} = 1|\text{data}) = .93$. The Bayes factor for inclusion can now be determined as follows (Huth, de Ron, et al., 2023; Marsman et al., 2022; Marsman & Haslbeck, 2023)

$$\underbrace{\frac{p(\text{data} \mid \gamma_{ij} = 1)}{p(\text{data} \mid \gamma_{ij} = 0)}}_{\substack{\text{Inclusion} \\ \text{Bayes factor}}} = \underbrace{\frac{p(\gamma_{ij} = 1 \mid \text{data})}{p(\gamma_{ij} = 0 \mid \text{data})}}_{\substack{\text{Posterior} \\ \text{inclusion odds}}} \Big/ \underbrace{\frac{p(\gamma_{ij} = 1)}{p(\gamma_{ij} = 0)}}_{\substack{\text{Prior} \\ \text{inclusion odds}}}.$$

The inclusion Bayes factor quantifies the weighted evidence for the inclusion of the relationship across all structures. As such, the inclusion Bayes factor provides a simple measure to distinguish between inconclusive evidence and conclusive conditional independence between two nodes. When we assume that all structures are equally likely *a priori*, the prior inclusion probability for individual edges is equal to 1/2. The prior odds then equal to 1, and we see that the inclusion Bayes factor for including the edge between variables A and I is equal to $.07/.93 \approx .074$, which means that based on the information in our data, we have strong evidence that an edge between variables A and I should be excluded from the network, in other words, we have strong evidence for conditional independence (i.e., the exclusion Bayes factor is $1/.074 \approx 13.5$). Note that the inclusion Bayes factor does not depend on the remaining relationships in the network, since it averages the network structures and thus overcomes the dependence of the single-model Bayes factor on assumptions about the remaining relationships.

Simulation study

We performed a simulation study to compare the accuracy of edge selection using the three methods in the case of a GGM using the `BDgraph` R package (Mohammadi & Wit, 2019). The R code we used in our simulations is available in the repository at https://osf.io/2x74v/. We simulated several conditions. Specifically, we varied the size of the network, $p = \{10,\ 30,\ 50\}$, the number of observations, $n = \{100,\ 200,\ 500,\ 1,000,\ 5,000\}$, the size of the focal edge weight between variables 1 and 2 (i.e., partial correlation), $\theta_{12} = \{0,\ .1,\ .25,\ .4\}$, and the density of the rest of the network (i.e., the number of relations in the rest of the network). We simulated the structures based on a random graph. We varied the density (D) of the network so that the probability of an edge between two nodes was either .2, 0.5, or .8. Given the generated structure, we sampled the remaining edge weights from a g-Wishart distribution (Roverato, 2002). Since manipulating the edge weight between variables 1 and 2 could result in a precision matrix that is not positive semi definite we continued sampling precision matrices until we found one that was positive semi definite.

We obtain the single-model (non-BMA) parameters, by sampling from their posterior distribution of the edge weights based on the full model. In this case this posterior distribution is a g-Wishart distribution (Lenkoski, 2013; Roverato, 2002). We obtain the single-

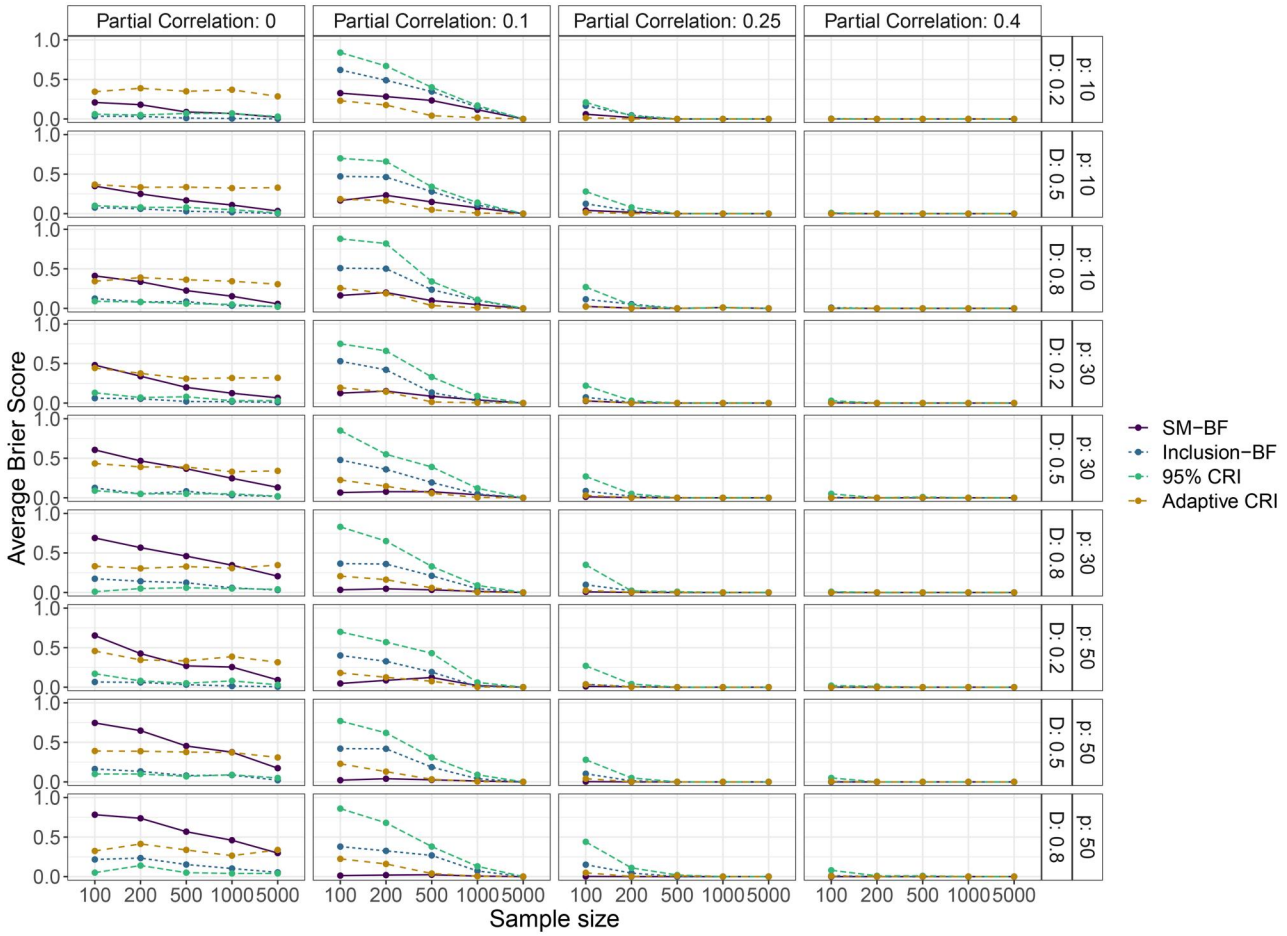

Figure 3. Average Brier score for each of the four measures as a function of the sample size plotted for each value of the edge weight, number of variables (*p*), and network density (D).

model Bayes factors by computing the fraction of the normalizing constants (i.e., marginal likelihoods) of the g-Wishart distributions under the fully connected structure and a structure that excludes the focal edge. For BMA analysis we used the default settings from the function `bdgraph`. For each dataset, using 10,000 iterations,[4] we computed for the focal edge weight θ_{12}:

1. The central credible interval for the single-model parameter estimate, and whether or not it included the test-relevant value of 0. We then transformed this into a quasi-inclusion probability, which was 1 if the interval included 0 and 0 otherwise, to make it comparable to the other measure. We computed two variants of the credible interval: (i) the standard 95% central credible interval and (ii) an adaptive credible interval. The latter is equivalent to ROPE (for more details, see, Kruschke, 2011).
2. The single-model posterior edge inclusion probability. The single-model posterior inclusion probability is calculated from the single-model Bayes factor as follows

$$p(\gamma_{ij} = 1|\text{data}) = \frac{\text{BF}_{10}\ \text{O}_{10}}{1 + \text{BF}_{10}\ \text{O}_{10}},$$

where BF_{10} is the single-model Bayes factor in favor of conditional dependence and O_{10} is the prior odds. We assumed full structure for the remaining relationships in the network and used $\text{O}_{10} = 1$ in our analysis.
3. The posterior edge inclusion probability obtained from Bayesian model averaging.

We computed the Brier score (Brier, 1950), which quantifies the mean squared difference between predicted probabilities and actual outcomes for a binary event (in this case the presence of an edge), with lower scores indicating better predictive performance.

[4]Note that we ran the MCMC procedures for a fixed number of iterations, and did not check for convergence of the individual Markov chains. Although our experience is that the implemented procedures tend to converge quickly, there is no guarantee that the chains that were used in our simulations actually did.

For each metric and condition Figure 3 shows that when the focal parameter has an edge weight equal to zero (i.e., conditional independence), the inclusion Bayes factor and the adaptive credible interval perform best across different sample sizes, numbers of variables, and network densities. We can also observe that the performance of the single-model Bayes factor becomes worse as the network density and the number of variables increase. All methods tend to perform better as a function of sample size. When the edge is present, even with a value of $\theta_{12} = 0.1$, we see that the situation is reversed, in other words, the 95% central credible interval and the single-model Bayes factor perform better than the inclusion Bayes factor and the adaptive credible interval, especially for $N < 1,000$. When the value of the partial correlation is 0.25 and 0.4, all of the methods tend to perform quite well.

Figure 3 shows that the density of the network has an influence on which method performs best. To get a clearer picture of the overall performance, we first aggregate the accuracy of the methods across effect sizes and compute the values for the area under the receiver operating characteristic curve (AUC) for each measure. The receiver operating characteristic (ROC) curve plots the tradeoff between the true positive rate (sensitivity) and the false positive rate (1 - specificity) as we vary the classification threshold. Therefore, the AUC is a performance measure of how well the methods can capture the truth—in this case, whether the edge is truly present. Methods with a higher AUC value (closer to 1) can better discriminate between present and absent edges than methods with lower AUC values (see Fawcett, 2006, for an introduction to ROC curves and AUC values). As can be seen from the results presented in Figure 4, the inclusion Bayes factor performs better than the single-model Bayes factor for low and medium network density levels, especially for smaller sample sizes, but performs worse when the network density is high. When the density is high, the structure assumed by the single-model Bayes factor is close to the true underlying network structure (i.e., both are densely connected), and thus the single-model Bayes factor has an advantage under this condition. The BMA approach still assumes different structures for the data and is therefore suboptimal when the true structure is dense. The 95% credible interval shows the worst performance overall.

Since the two Bayes factor approaches are the only formal ways to test for conditional independence hypotheses, we wish to compare their performance in some more detail. Figure 5 plots the proportion of times the Bayes factors made a correct decision in

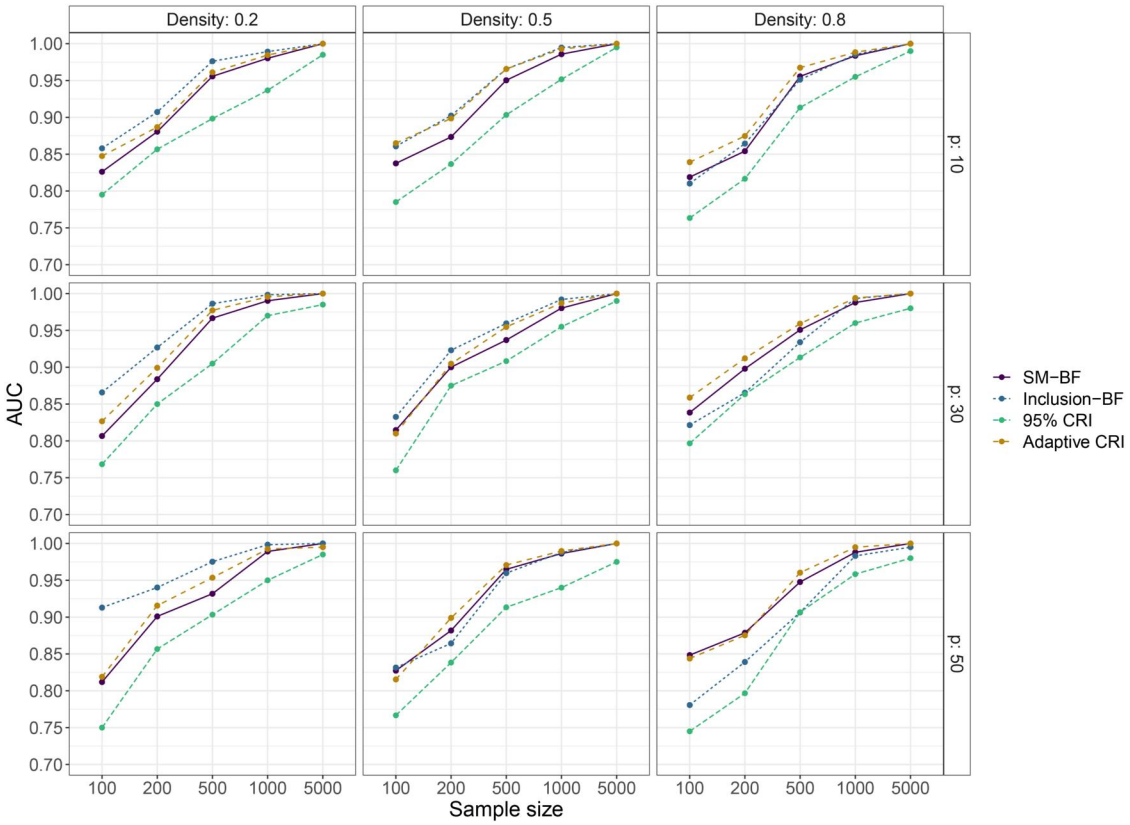

Figure 4. AUC values as a function of the sample size plotted for different values of the network density and number of variables *p*.

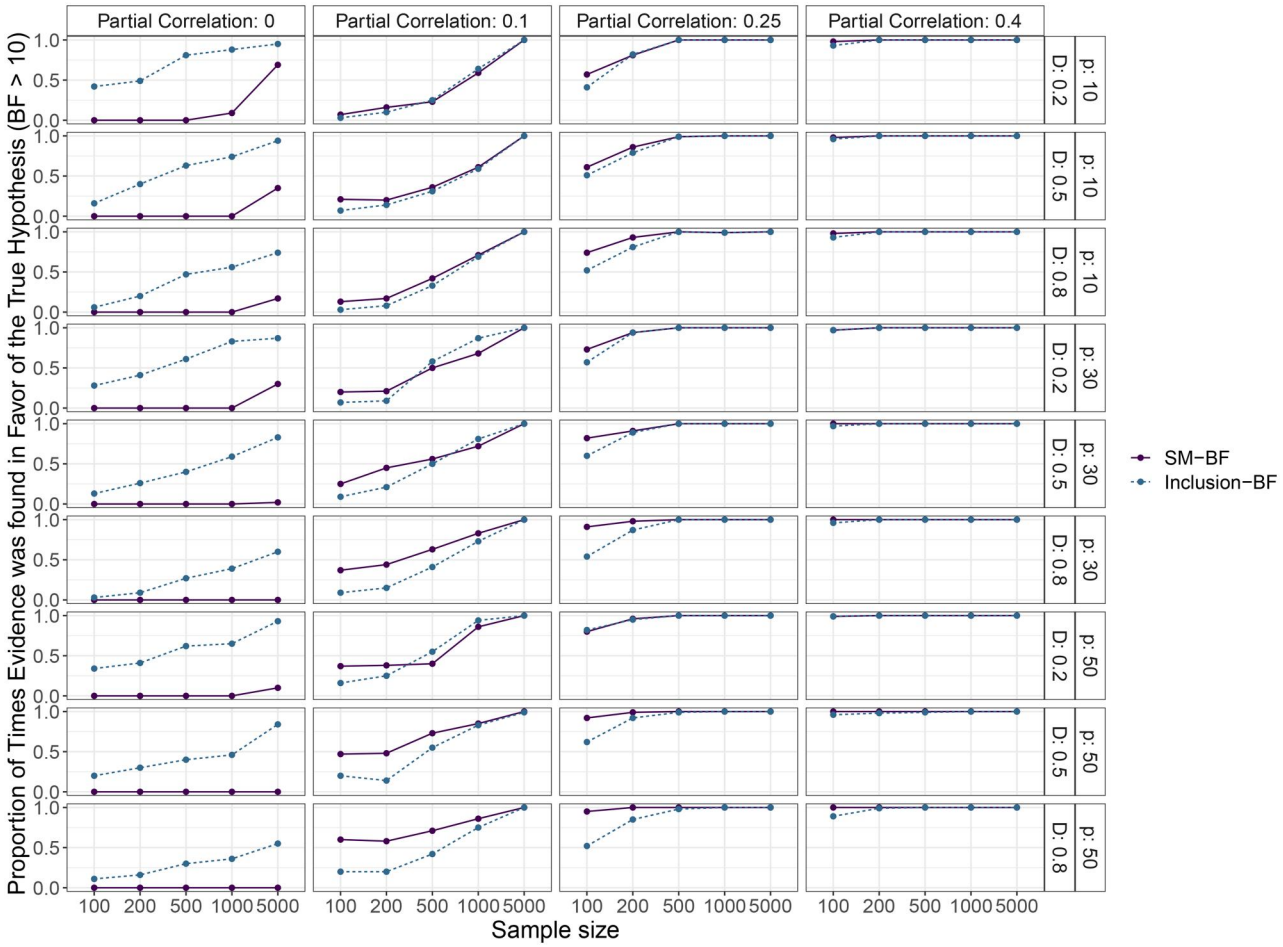

Figure 5. The proportion of times that the two Bayes factors found evidence in favor of the true hypothesis, as a function of the sample size plotted for each value of the edge weight, number of variables (*p*), and network density (D).

detecting evidence for the true hypothesis. As can be seen, and as expected based on the previous plots, when the edge is absent (i.e., when $\theta_{12} = 0$), the inclusion Bayes factor outperforms the single-model Bayes factor in all simulation conditions. This suggests that the inclusion Bayes factor is quite good at capturing evidence in favor of conditional independence. When the edge is present, its weight is small, and the network is small and sparse, the two Bayes factors show similar performance. In contrast, as can also be seen in Figure 4, as the network becomes larger and more densely connected, the single-model Bayes factor begins to outperform the inclusion Bayes factor. As the true value for the edge weight increases, both methods perform very well, especially for large sample sizes.

Empirical example

To illustrate the difference between the two Bayes factors we consider the analysis of a data set from a study by Gojković et al. (2022) on the network structure of empathy, narcissism, and the Dark Triad (i.e., the combination of narcissism, psychopathy, and Machiavellianism) personality traits. The data are publicly available at https://osf.io/7jcks/. It consists of eight variables, each measured by a battery of Likert-scale items. The narcissism, psychopathy, and Machiavellianism variables are based on the 27 items from the Short Dark Triad (i.e, each variable is a sum of responses to 9 different items Jones & Paulhus, 2014); the cognitive empathy, affective resonance, and affective dissonance variables are based on the 36 items from the Affective and Cognitive Measure of Empathy (Vachon & Lynam, 2016); the narcissistic admiration and narcissistic rivalry variables are based on the 18 items from the Narcissistic admiration (Adm)iration and Narcissistic Rivalry (Back et al., 2013). The affective dissonance items were inversely coded so that a higher summed score corresponded to a higher level of affective dissonance. The study was based on a sample of 263 high school and university students from Vojvodina, Serbia.

Since we wish to see if there is a difference in the conclusion we would draw from using the two Bayes

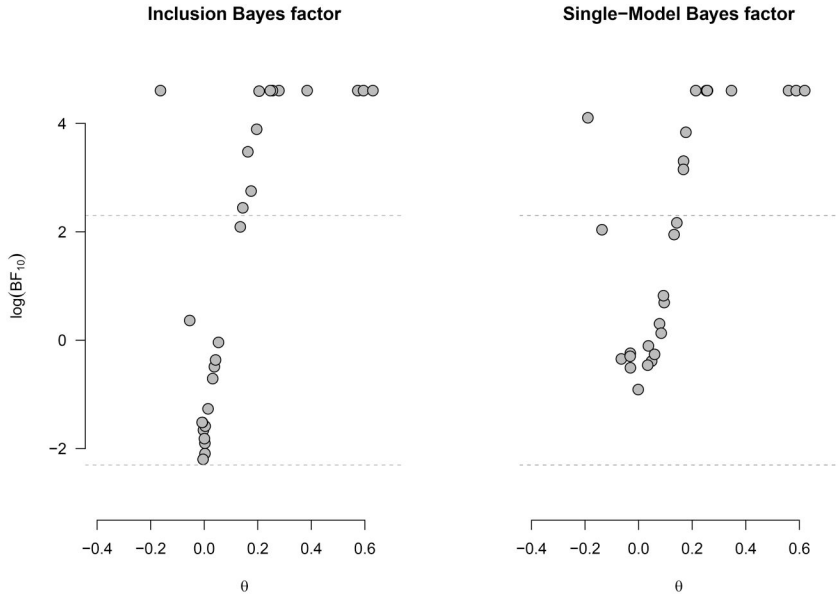

Figure 6. The (natural) logarithm of the Bayes factors plotted against the posterior mean of the corresponding edge weight. The left panel shows the results for the inclusion Bayes factor, and the right panel shows the results for the single-model Bayes factor. Bayes factor values greater than or equal to one hundred are set equal to one hundred (i.e., $\log(BF_{10}) = 4.6$).

factors, we analyzed the network structure of the eight variables with a GGM using both the single-model and multi-model or BMA perspectives. For the single-model analysis, we estimated the parameters of a fully connected GGM by drawing one million samples from the corresponding posterior distribution, which in this case is a g-Wishart distribution (Lenkoski, 2013; Roverato, 2002). The Bayes factor was computed for each of the $8 \times (8-1)/2 = 28$ edges in the network by computing the ratio of the marginal likelihood with all edges present to the marginal likelihood with the focal relationship excluded. We used the `BDgraph` package to sample from the g-Wishart distribution and to compute the marginal likelihood. For the multi-model analysis, we also used the `BDgraph` package, which estimates the posterior inclusion probabilities using a Markov chain Monte Carlo procedure. We used one million iterations for each Markov chain. In each of these analyses, we used the default settings of `BDgraph`, setting a g-Wishart prior on the precision matrix $\boldsymbol{\Theta}$ and assuming a prior inclusion probability of 1/2 for all edges.

Figure 6 illustrates that there is indeed a difference between the inferences we would draw using the inclusion Bayes factor and the single-model Bayes factor. We can see that the inclusion Bayes factor provides evidence for edge exclusion, i.e., for the estimated parameters that are close to zero, as indicated by the narrower" v" shape shown in the left panel. In the BMA case, there is a more pronounced shrinkage toward zero. Therefore, as shown in the previous section, the inclusion Bayes factor offers more pronounced evidence in support of conditional independence than the single-model Bayes factor.

Figure 7 shows the edge evidence plots—networks whose edges reflect strong evidence for edge inclusion (using a cutoff of $BF_{10} = 10$). Based on the inclusion Bayes factor in the left panel, we conclude that 13 of the 28 possible edges are present in the network, and based on the single-model Bayes factor in the right panel, we conclude that 12 of them are present. For the edge between the variables psychopathy (SD3P) and affective resonance (ARe), the exclusion Bayes factor is equal to $BF_{01} = 9.1$, close to the evidential cutoff of 10, giving us evidence in favor of conditional independence. For comparison, the largest single-model Bayes factor in favor of edge exclusion is between the variables admiration (Adm) and affective resonance (ARe), and is only $BF_{01} = 2.5$. Examining the networks in Figure 7, we can see that, for example, with the inclusion Bayes factor we find evidence for the inclusion of an edge between the variables psychopathy (SD3P) and admiration (SD3N), cognitive empathy (CEm) and admiration (SD3N), but we have no evidence for the inclusion of the same edges when we use the single-model Bayes factor. Conversely, using the single-model Bayes factor, we find evidence for the inclusion of an edge between the variables cognitive empathy (CEm) and admiration (Adm), for which we have inconclusive evidence when using the inclusion Bayes factor. From our simulations, we know that When the network structure is

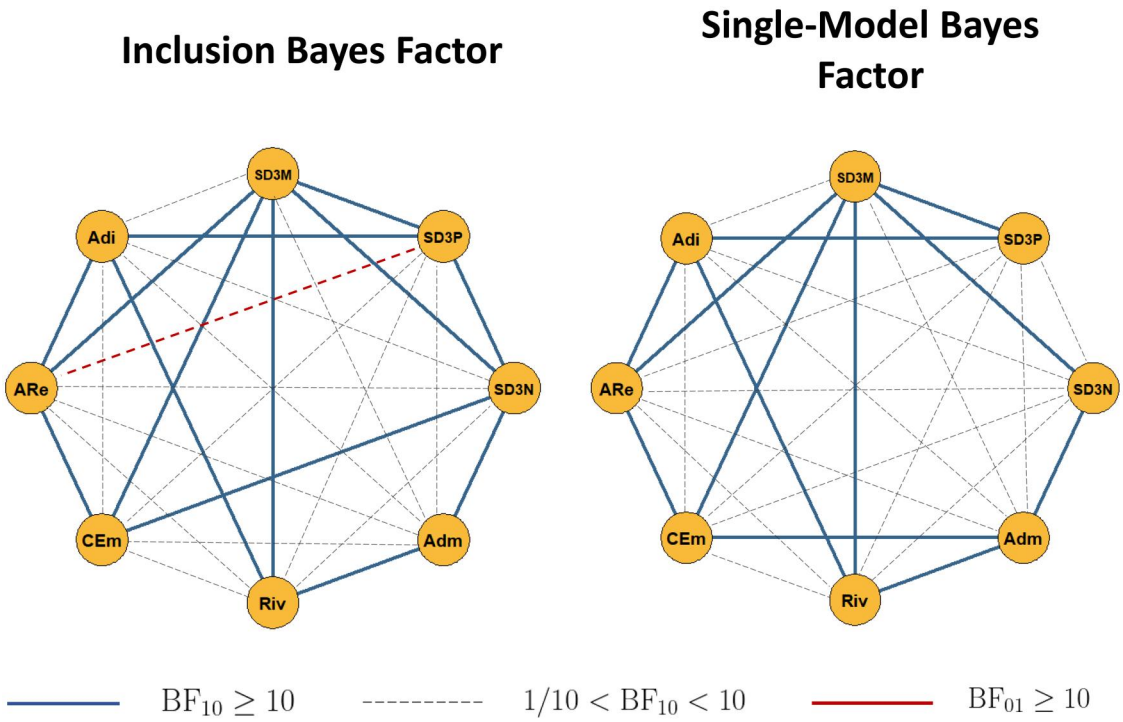

Figure 7. Edge evidence plots based on the inclusion Bayes factor on the left and the single-model Bayes factor on the right. The blue solid lines indicate edges for which there is a $BF_{10} \geq 10$, the dashed red line indicates an inclusion Bayes factor that almost reaches the exclusion threshold and the dashed grey lines indicate edges for which there is inconclusive evidence for edge (in)exclusion.

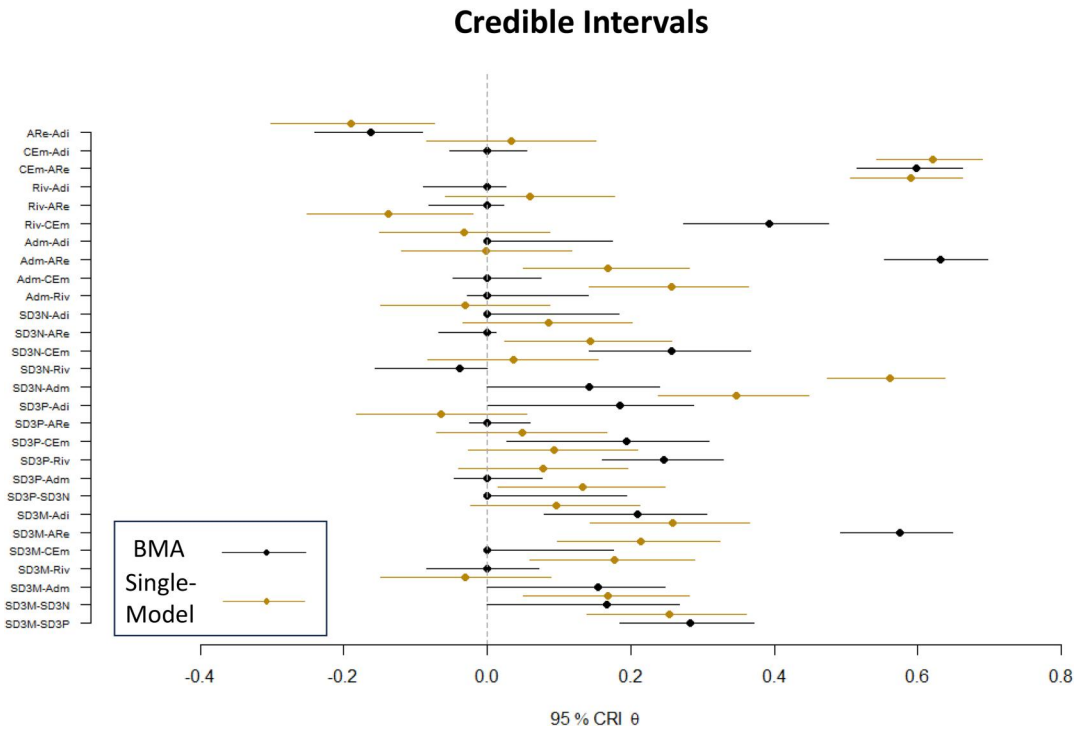

Figure 8. The 95% credible intervals for the BMA estimates in black and for the estimates based on a fully connected structure in yellow. The vertical dotted line represents the value of $\theta = 0$. The points on each line represent the posterior median estimates.

sparse, which appears to be the case in this example, the inclusion Bayes factor can more accurately capture the evidence, both for edge inclusion and edge exclusion.

Since we argue that the credible and/or highest density intervals should not be used for hypothesis testing, we adhere to this principle in this section. However, because these intervals are valuable measures of posterior parameter uncertainty, we present plots of the 95% central credible intervals around each posterior edge weight. We computed the 95% central credible intervals for the BMA parameter estimates and the 95% central credible intervals for the posterior parameter estimates based on a structure that assumes all edges are present. As can be seen in Figure 8, the central credible intervals obtained from the two methods are different. We prefer the credible intervals based on BMA because they account for both parameter uncertainty and structure uncertainty.

Discussion

In this paper, we have reviewed three different Bayesian approaches to testing conditional independence hypotheses for a class of Markov random field models used in network psychometrics. The first method uses the posterior distribution of the partial association θ_{ij} to check whether it falls in the ROPE, or similarly whether its $x\%$ credible interval contains zero. Both scenarios would indicate that the hypothesis of conditional independence of the variables i and j cannot be rejected, but the drawback is that we cannot use it to support the independence hypothesis. The second approach used the single-model Bayes factor to test for conditional independence, which compares two network structures $\mathcal{S}_s$ and $\mathcal{S}_t$ that are identical except that the focal relationship is included in $\mathcal{S}_s$ but not in $\mathcal{S}_t$. Although this method could be used to express support for the conditional independence hypothesis, its drawback is that it is sensitive to the required choice of which relations are in the rest of the network. The third approach uses BMA to express the inclusion Bayes factor, which accounts for the uncertainty about other relations in the network. The inclusion Bayes factor is free from the conceptual problems of credible interval-based tests and is optimal when we are uncertain about the structure underlying our data.

In the simulations, we showed that the inclusion Bayes factor was the best overall method for determining conditional independence. It also performed well in determining conditional dependence, although the single-model Bayes factor outperformed the inclusion Bayes factor in scenarios where the true network structure is densely connected. In these scenarios, which are close to the assumption of a fully connected structure underlying the single model Bayes factor, the inclusion Bayes factor loses power because it continues to consider alternative structures for the data at hand. However, in practice, since we do not know what the underlying structure is, the inclusion Bayes factor is the most robust choice for inferring conditional independence or dependence.

Critique: The correct model is probably not being considered

The mathematics behind Bayesian model comparison does not assume that any of the models under consideration are correct in some abstract sense, as the formulas only evaluate the predictive adequacy of the models under consideration (see for instance, O'Hagan, 2010, p. 167). Nevertheless, many statisticians have argued that Bayesian model comparison only makes sense if the correct model is in the collection of models under consideration—the $\mathcal{M}$-closed context (Bernardo & Smith, 1994, pp. 383–407). The main concern of critics of Bayesian model comparison, and BMA in particular, is that the posterior distribution cannot converge to the correct model if it is not in the collection of models under consideration—the $\mathcal{M}$-open context. Instead of converging to the correct model, the posterior distribution would converge to the model that is closest to the true model in a Kullback-Leibler sense in the $\mathcal{M}$-open context. This model would be optimal in terms of its predictive adequacy relative to the collection of models under consideration.

Box's famous adage "all models are wrong" (Box, 1976, p. 792) is often used to make the case that the $\mathcal{M}$-closed assumption is also wrong. There are two ways in which we think the true model might differ from the one we consider in psychological network modeling. First, the network models we use typically include main effects and pairwise relations (i.e., first and second-order interactions). In principle, one could consider models with third or higher-order interactions, but these models are computationally demanding. Second, we often have a substantive motivation for choosing the variables to include in our network, but this choice can have a huge impact on the network structure. For example, two variables will be conditionally dependent if we exclude their common cause from the network, but conditionally

independent if we include it. This is called the boundary specification problem (Laumann et al., 1989; Neal & Neal, 2023). However, it is likely that if we knew which variable caused other variables, we would include it in the network. Thus, while we agree that the $\mathcal{M}$-closed assumption is unlikely to hold in practice, we also agree with the continuation of the adage that "all models are wrong, but some models are useful" (Box & Draper, 1987, p. 424). With BMA, we evaluate the predictive adequacy of the structures of interpretable network models formulated on a substantively interesting subset of variables.

Limitation: There are no substantively motivated or good default prior distributions for psychological networks

The BMA approach requires us to specify our prior knowledge and expectations about the structure of the network. However, despite the large body of literature on psychological network modeling, we still have a limited understanding of their structure. The main reason for this limited understanding is that Bayes factor tests that can quantify the support for certain relational patterns have only recently been proposed and have not yet gained much traction. As a result, we must rely on standard, objective specifications of priors for psychometric topologies and their associated parameters. These priors may be inappropriate for psychological networks for two reasons. First, they may give relatively little weight to the correct structure. In objective specifications, we usually assign a uniform prior on the possible structures (cf. appendix). Since there is often a huge collection of possible structures for the network under consideration, only a tiny fraction of the total probability is assigned to the correct model. Thus, finding the right model is like looking for a needle in a haystack. Therefore, it would be helpful to know in advance what kind of model we are looking for. Second, objective priors are often developed in the context of regression models, and we are unsure if these specifications make sense in the network context. For example, in regression, it makes sense to find a sparse collection of variables that can make accurate predictions. This is because we wish to choose the least complex model (i.e., the model with the fewest number of predictors) that best predicts new data. But, in the context of MRF models an absent edge carries a strong assumption, namely of conditional independence, which indicates that we should not exclude edges by default. Although the objective priors we use here assign equal probability to including and excluding individual edges, we need to investigate the suitability of these priors in the network context. We encourage researchers to always perform sensitivity analyses by estimating the models under different prior specifications and examining whether and how much the different specifications alter the conclusions.

In order to advance the specification of good prior densities, we need to advance our understanding of psychometric network structures. Early discussions about the underlying structures of psychometric networks were a reaction to the massive popularity of lasso-based methods, which assume sparse network structures. Alternatives to the lasso have been proposed that either focus on densely connected networks (e.g., Marsman et al., 2015), or that aim to strike a balance between sparse and dense network topologies (e.g., Chen et al., 2018), but these approaches have not been widely adopted. This means that we must interpret the sparsity of psychometric networks with caution, especially when data are limited (Epskamp et al., 2017; Williams et al., 2019).

Now that BMA allows us to test our predictions about network topology, we are entering a new era of network psychometrics. In the next decade, armed with new Bayesian methodology, we hope to see an advanced understanding of the structure of psychometric networks, how they differ across measures and populations, and which relationships have been explained and which have not.

Limitation: There are few BMA methods for analyzing psychological networks

For network researchers to adopt BMA for their analyses, it is imperative that the methodology be implemented in user-friendly software. Most psychological network modeling analyses are performed in the statistical software R, and two R packages now implement BMA for network analysis. The `BDgraph` package[5] includes methods for analyzing continuous, binary, and ordinal variables (GGMs and latent GGMs; R. Mohammadi & Wit, 2019), and the `bgms` package[6] for analyzing MRFs of (mixed) binary and ordinal variables (Marsman & Haslbeck, 2023). Since most data sets in psychology contain binary and ordinal variables, these two R packages already cover a lot of ground. The `BDgraph` package is now also implemented in the open-source statistical software JASP (see Huth, de Ron, et al., 2023), which has a graphical

[5]https://cran.r-project.org/web/packages/BDgraph/index.html.
[6]https://cran.r-project.org/web/packages/bgms/index.html.

user interface that allows users to point and click on their desired analyses (e.g., Love et al., 2019; Wagenmakers, Love, et al., 2018). The JASP implementation opens BMA-based methods for psychological networks to researchers without experience programming in R.

Although we argue that Bayes factor approaches, especially the inclusion Bayes factor, should be preferred for testing conditional independence hypotheses, as shown in the empirical example, one can still use the credible or highest density intervals around the model-averaged parameter estimates (e.g., edge weights) as measures of parameter uncertainty. For these and many other advantages, the interested reader is referred to the newly developed R package `easybgm` (Huth, Keetelaar, et al., 2023), which allows researchers with less programming experience to use powerful packages such as `bgms` and `BDgraph` to analyze their data and obtain (BMA) Bayes factors as well as edge uncertainty plots.

The existing software for BMA-based methods for the analysis of psychological networks covers several important variable types—e.g., continuous, binary, and ordinal variables—for cross-sectional applications of networks. However, there are currently no software solutions for networks with nominal, discrete, or count variables, or for longitudinal data designs. The development of BMA methods for analyzing these types of variables and research designs, and their software implementations, is a fruitful area for future research.

Challenge: Bayesian model averaging can be time consuming

One of the main challenges of BMA is that it must evaluate the collection of models under consideration. In practice, it is rarely possible to enumerate all possible models, since the number of structures grows rapidly as the number of variables increases. Therefore, the R packages that estimate these models rely on Stochastic Search Variable Selection techniques (George & McCulloch, 1993). These techniques are typically implemented through Markov chain Monte Carlo algorithms (MCMC, see van Ravenzwaaij et al., 2018, for an accessible introduction) that iteratively simulate a network structure and its associated parameters from the joint posterior distribution. As mentioned in the section on prior distributions, first an edge indicator variable γ_{ij} is sampled, and then the corresponding edge weight θ_{ij} is assigned to a particular prior distribution given the sampled value for the edge indicator. Since the space of possible models is usually large, it is imperative to run such procedures for enough iterations to sufficiently explore the joint posterior distribution. For some models, such as the GGM, this is usually very fast for the size of data sets encountered in psychology. However, for binary or ordinal models, MCMC procedures can take a long time, depending on the sample size. Fortunately, we only need to run the procedure once to get the full Bayesian benefit.

Conclusion

We have provided a conceptual review of recent Bayesian tests for conditional independence of variables in psychological networks. We argued that the two Bayes factor tests are conceptually superior to frequentist and credible interval-based tests for conditional independence, in particular because they can express support, or lack thereof, for conditional independence and dependence between the network's variables. We have shown that the single-model Bayes factor is sensitive to the assumption that must be made about the underlying network structure, while the inclusion Bayes factor adequately accounts for the structure uncertainty. Thus, the inclusion Bayes factor provides researchers with a straightforward test of conditional independence and dependence hypotheses. We hope that the new Bayesian methodology, which focuses on the analysis of the structure of psychological networks, (i.e., psychometric topology) will help unravel the complex systems underlying psychological variables.

Article Information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: NS, SEK, and MM were supported by the European Union [ERC, BAYESIAN P-NETS, #101040876]. 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. KH was supported by the Center for Urban Mental Health (University of Amsterdam)

and DvdB was supported by Amsterdam Brain and Cognition (University of Amsterdam).

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Back, M. D., Küfner, A. C., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870–897. https://doi.org/10.1214/009053604000000238

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*(3), 317–335. https://doi.org/10.1214/ss/1177013238

Berger, J. O., & Pericchi, L. R. (2015). Bayes factors. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online*. Wiley. https://doi.org/10.1002/9781118445112.stat00224.pub2

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory*. Wiley.

Blanken, T. F., Isvoranu, A.-M., & Epskamp, S. (2022). Estimating network structures using model selection. In *Network Psychometrics with R* (pp. 111–132). Routledge.

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., van Borkulo, C. D., van Bork, R., & 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

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons, Inc.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika*, *83*(3), 538–562. https://doi.org/10.1007/s11336-018-9610-4

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

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, *88*(2), 71–83. https://doi.org/10.1159/000497425

Dienes, Z. (2014). Using Bayes to get the most our of non-significant results. *Frontiers in Psychology*, *5*(781), 781. https://doi.org/10.3389/fpsyg.2014.00781

Dobra, A., & Lenkoski, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *The Annals of Applied Statistics*, *5*(2A), 969–993. https://doi.org/10.1214/10-AOAS397

Eberhardt, F. (2017). Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics*, *3*(2), 81–91. https://doi.org/10.1007/s41060-016-0038-6

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

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., Haslbeck, J. M. B., Isvoranu, A. M., & van Borkulo, C. D. (2022). Pairwise Markov random fields. In A. M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network psychometrics with R: A guide for behavioral and social scientists* (pp. 93–110). Routledge, Taylor & Francis Group.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One*, *12*(6), e0179891. https://doi.org/10.1371/journal.pone.0179891

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

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/10.1080/01621459.1993.10476353

Giudici, P. (1995). Bayes factors for zero partial covariances. *Journal of Statistical Planning and Inference*, *46*(2), 161–174. https://doi.org/10.1016/0378-3758(94)00101-Z

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

Gojković, V., Dostanić, J. S., & Đurić, V. (2022). Structure of darkness: The dark triad, the 'dark' empathy and the 'dark' narcissism. *Primenjena Psihologija*, *15*(2), 237–268. https://doi.org/10.19090/pp.v15i2.2380

Good, I. J. (1985). Weight of evidence: A brief survey. *Bayesian Statistics*, *2*, 249–270.

Guo, J., Levina, E., Michailidis, G., & Zhu, J. (2015). Graphical models for ordinal data. *Journal of Computational and Graphical Statistics: A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, *24*(1), 183–204. https://doi.org/10.1080/10618600.2014.889023

Haslbeck, J. M. B., & Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, *93*(8), 1–46. https://doi.org/10.18637/jss.v093.i08

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215. https://doi.org/10.1177/2515245919898657

Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–401. https://doi.org/10.1214/ss/1009212519

Huth, K. B. S., de Ron, J., Goudriaan, A. E., Luigjes, J., Mohammadi, R., van Holst, R. J., Wagenmakers, E.-J., & Marsman, M. (2023). Bayesian analysis of cross-sectional networks: A tutorial in R and JASP. *Advances in Methods and Practices in Psychological Science*, *6*(4), 193. https://doi.org/10.1177/25152459231193

Huth, K., Keetelaar, S., Sekulovski, N., van den Bergh, D., & Marsman, M. (2023). Simplifying Bayesian analysis of graphical models for the social sciences with easybgm: A user-friendly R-package. PsyArXiv. https://doi.org/10.31234/osf.io/8f72p

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift Für Physik*, *31*(1), 253–258. https://doi.org/10.1007/BF02980577

Jeffreys, H. (1939). *Theory of probability*. Clarendon Press.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (sd3) a brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Jongerling, J., Epskamp, E., & Williams, D. R. (2023). Bayesian uncertainty estimation for Gaussian graphical models and centrality indices. *Multivariate Behavioral Research*, *58*(2), 311–339. https://doi.org/10.1080/00273171.2021.1978054

Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. *Psychometrika*, *86*(1), 215–238. https://doi.org/10.1007/s11336-021-09754-5

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

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayesian factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kindermann, R., & Snell, J. L. (1980). *Markov Random Fields and their Applications*. (Vol. 1) American Mathematical Society.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *6*(3), 299–312. https://doi.org/10.1177/1745691611406925

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.

Lenkoski, A. (2013). A direct sampler for G-Wishart variates. *Stat*, *2*(1), 119–128. https://doi.org/10.1002/sta4.23

Lindley, D. (2004). That wretched prior. *Significance*, *1*(2), 85–87. https://doi.org/10.1111/j.1740-9713.2004.026.x

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP – Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

Marsman, M. (2023). Bgms: Bayesian variable selection for networks of binary and/or ordinal variables [Computer software manual]. (R package version 0.1.0)

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., Maas, H. L. J. 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., & Haslbeck, J. M. B. (2023). Bayesian analysis of the ordinal Markov random field. PsyArXiv. https://osf.io/preprints/psyarxiv/ukwrf

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*(1), 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

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), 3. https://doi.org/10.18637/jss.v089.i03

Morey, R. D., Hoekstra, R. H. A., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

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. (2010). *Kendall's Advanced Theory of Statistic 2B*. John Wiley & Sons.

Open Science Foundation. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 716. https://doi.org/10.1126/science.aac4716

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Robinaugh, D. J., Hoekstra, R. H. A., 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

Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics*, *29*(3), 391–411. https://doi.org/10.1111/1467-9469.00297

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*, *29*(6), 953–970. https://doi.org/10.1080/10705511.2022.2056039

Sekulovski, N., Keetelaar, S., Haslbeck, J. M. B., Marsman, M. (2023). Sensitivity analysis of prior distributions in bayesian graphical modeling: Guiding informed prior choices for conditional independence testing. *PsyArXiv*. https://doi.org/10.31234/osf.io/6m7ca

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). MIT Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135–149. https://doi.org/10.1177/1073191114567941

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. *Scientific Reports*, *4*(1), 5918. (https://doi.org/10.1038/srep05918

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842–860. https://doi.org/10.1111/cdev.12169

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Vanpaemel, W., & Lee, M. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*(6), 1047–1056. https://doi.org/10.3758/s13423-012-0300-4

van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, *25*(1), 143–154. https://doi.org/10.3758/s13423-016-1015-8

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Lee, M. D., Rouder, J. N., & Morey, R. D. (2020). The principle of predictive irrelevance or why intervals should not be used for model comparison featuring a point null hypothesis. In C. W. Gruber (Ed.), *The theory of statistics in psychology – Applications, use and misunderstandings*. Springer.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E.-J., 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 Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, *25*(3), 169–176. https://doi.org/10.1177/0963721416643289

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they Analyze their data: The case of psi:

Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Waldorp, L. J., & Marsman, M. (2022). Relations between networks, regression, partial correlations, and the latent variable model. *Multivariate Behavioral Research*, *57*(6), 994–1006. https://doi.org/10.1080/00273171.2021.1938959

Williams, D. R. (2021). 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., & 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

Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, *54*(5), 719–750. https://doi.org/10.1080/00273171.2019.1575716

Appendix. Prior distributions for MRF models implemented in R packages

The Bayesian analysis of an MRF model requires specifying two sets of prior distributions.

Priors on the structure

The prior probabilities on the network structure $p(\mathcal{S}_s)$ can be expressed by specifying a prior probability on the possible value for each binary indicator variable γ_{ij}. This is achieved by assuming that each edge follows an independent Bernoulli distribution with a prior inclusion probability π_{ij}. The R packages `bgms` (Marsman, 2023) for analyzing MRF for binary and ordinal data and `BDgraph` (R. Mohammadi & Wit, 2019) for analyzing GGMs both provide this as the default option for the prior on the network structure. Setting this prior with $\pi_{ij} = 0.5$ for all edges is considered an uninformative or objective choice, and this option is also referred to as the uniform prior on the structure (e.g., Marsman et al., 2022). Of course, there are other prior options for the network structure that take into account the number of present edges (i.e, the complexity of the structure); we refer the interested reader to Huth, de Ron, et al. (2023) for an accessible introduction to these priors and to Sekulovski, Keetelaar, Haslbeck, and Marsman (2023) for a detailed discussion of prior selection with particular emphasis on the priors implemented in the R package `bgms`.

Priors on the edge weights

We also need to specify priors on the edge weight parameters in $\boldsymbol{\Theta}$. The R package `BDgraph` specifies a g-Wishart distribution (Roverato, 2002) on the precision matrix (i.e., the inverse of the covariance matrix) containing the (untransformed) edge weight parameters for the GGM. The g-Wishart distribution takes two parameters (i) the degrees of freedom d, which is by default set to $d = 3$, and (ii) a scale matrix $\mathbf{D}$, set by default to an uninformative $p \times p$ identity matrix. The R package `BGGM` (Williams & Mulder, 2020b), which can also be used to analyze GGMs, specifies either a Wishart or a Matrix F prior on the precision matrix (Williams & Mulder, 2020a). The (g-)Wishart priors are conjugate to the precision matrix and assure a posterior density function on the space of positive semi definite matrices. Note that the R package `BGGM` does not stipulate priors on the network structure since it assumes that all edges are present a priori. Finally, the R package `bgms` specifies prior distributions on the individual edge weights given the value of the edge indicator variable γ_{ij}, i.e., $p(\theta_{ij}|\gamma_{ij})$. In other words, if $\gamma_{ij} = 0$, the edge weight is set to zero, and if $\gamma_{ij} = 1$, the edge is given a specific (diffuse) prior distribution (e.g., a Cauchy distribution). For more details, see Marsman and Haslbeck (2023) and Sekulovski et al. (2023).