

Structural and parametric uncertainties in full Bayesian and graphical lasso based approaches: beyond edge weights in psychological networks

Gabor Hullam^{*,†}, Gabriella Juhasz^{†,‡,§}, Bill Deakin[§] and Peter Antal^{*}

^{*}Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest Hungary
[†]MTA-SE Neuropsychopharmacology and Neurochemistry Research Group, Hungarian Academy of Sciences, Budapest, Hungary

[‡]Department of Pharmacodynamics, Faculty of Pharmacy, Semmelweis University, Budapest, Hungary

[§]Neuroscience and Psychiatry Unit, School of Community Based Medicine, Faculty of Medical and Human Sciences, The University of Manchester, Manchester, UK

Abstract—Uncertainty over model structures poses a challenge for many approaches exploring effect strength parameters at system-level. Monte Carlo methods for full Bayesian model averaging over model structures require considerable computational resources, whereas bootstrapped graphical lasso and its approximations offer scalable alternatives with lower complexity. Although the computational efficiency of graphical lasso based approaches has prompted growing number of applications, the restrictive assumptions of this approach are frequently ignored, such as its lack of coping with interactions. We demonstrate using an artificial and a real-world example that full Bayesian averaging using Bayesian networks provides detailed estimates through posterior distributions for structural and parametric uncertainties and it is a feasible alternative, which is routinely applicable in mid-sized biomedical problems with hundreds of variables. We compare Bayesian estimates with corresponding frequentist quantities from bootstrapped graphical lasso using pairwise Markov Random Fields, discussing also their interpretational differences. We present results using synthetic data from an artificial model and using the UK Biobank data set to explore a psychopathological network centered around depression (this research has been conducted using the UK Biobank Resource under Application Number 1602).

I. INTRODUCTION

The joint exploration of the existence and quantitative properties of all direct dependency relations in a given domain, i.e. which are not mediated by other variables present in the given analysis, is still a central challenge. At the two ends of the spectrum of approaches utilizing recent large health data sets are the (1) knowledge-intensive parametric methods and the (2) fully inductive methods. The former relies on creating an *a priori* network based on a preliminary hypothesis, and using data for its validation and refinement. The aim of such methods is to create an adequate parametric model that allows the investigation of quantitative effect strength relations. The latter approach uses the data in inductive methods potentially incorporating *a priori* constraints to investigate quantitatively the effect strength relations.

The growing availability of large data sets boosted expectations that inductive methods can cope with the model structure level as well, but theoretical results indicated that

computational and statistical complexities are prohibitive in general. Within the frequentist approaches, severe restrictions of model classes, sparsity-oriented complexity regularization, and utilization of bootstrapping led to tractable methods, such as graphical lasso and its extensions. Within the Bayesian framework, abundant computational resources fuel the development of Bayesian averaging methods over model structures to bypass the problem of model structure identification.

The low computational costs and robust statistical properties of graphical lasso prompted widespread applications even in domains, where its assumptions about the underlying distribution, high-dimensionality, scarcity of samples and lack of interactions are not satisfied or not that vital. The application of the bootstrap framework for structural model properties poses further questions, which led to the following two questions investigated in this paper:

- 1) **Effect strength parameters:** What is the relation of effect strength estimates from graphical lasso, from bootstrapped graphical lasso and from a full Bayesian averaging over model structures and parameters?
- 2) **Relational existence:** What is the relation of bootstrap proportions and posteriors for structural model properties, e.g. edges?

We systematically investigate these correspondences using a synthetic data set from an artificial model and a depression-related subset of the UKBiobank data set, discussing also novel results from this analysis.

In the paper, first we overview properties of graphical lasso based and full Bayesian approaches for the joint exploration of parametric and structural aspects of direct dependencies. Next, we briefly overview a specific full Bayesian approach, which uses Markov-blanket graphs from the off-line results of an MCMC analysis. Next, we systematically investigate the effect of sample size on graphical lasso based estimates and on full Bayesian estimates using an artificial network containing 11 variables. Finally, we systematically compare graphical lasso based estimates and full Bayesian estimates using the UK

Biobank data set to explore a depression focused comorbidity and deep phenotypic network.

II. PREVIOUS WORKS

The knowledge-intensive, probabilistic graphical model (PGM) related approaches towards exploring the existential and parametric aspects of direct relations usually require a predefined structure which is evaluated and potentially refined against the data. Related methods, such as the frequently applied structural equation modeling, provide measures for assessing the significance of variable relationships and model fitting [1]. Furthermore, the prequential approach allowed a sequential Bayesian evaluation of various, hierarchic model properties [2]. These approaches can be advantageous for relatively small, well-defined models based on firm hypotheses or former evidence. However, the lack of learning the structure from data poses considerable limitations e.g. in case of previously unexplored domains.

Inductive methods use data (usually without flexible options to incorporate a priori knowledge) to jointly explore both the existential and parametric aspects of direct relations. Early frequentist methods, such as the causally motivated TETRAD approach were designed for the frequentist identification of structure and parameters, with proven asymptotic performance under its assumptions [3]. Recent frequentist PGM-related methods, such as graphical lasso, focus on the learning of sparse models to allow applicability for small relative sample size and for high-dimensional data sets, even at the omic scales [4], [5]. Full Bayesian approaches usually take advantage of the existence of efficiently computable, closed form analytic solutions for model averaging at the parameter level and the availability of exact Bayesian averaging in smaller problems [6], [7] and the availability of efficient Markov Chain Monte Carlo (MCMC) schemes for performing Bayesian averaging at the model structure level [8].

Because of the growing interest of using bootstrapped graphical lasso in such mid-sized problems, where MCMC-based, hybrid full Bayesian approaches are also available and can be used as a reference, we provide an overview on these two approaches, and we investigate two specific methods.

A. Graphical lasso based approaches

The research of probabilistic graphical models as sparse representations for correlation (covariance) structures goes back to the work of Dempster [9]. This line of research was closely related to the investigation of the representational capabilities of probabilistic graphical models (PGMs), using undirected graphs (Markov networks, Markov Random Fields, MRFs) and directed acyclic graphs (Bayesian networks, BNs) [10]–[13]. The first lasso proposals to learn sparse probabilistic graphical models (“graphical lasso”) assumed multivariate normal distributions [4], [5], which were extended toward the Bayesian framework [14], [15] and also toward non-Gaussian, discrete cases [16], [17]. Extensions for binary data resulting in binary pairwise Markov Random Fields (bPRMs), were also reported based on predictive approximations [18], [19].

A central, simplifying assumption behind this approach is the assumption of linear Gaussian relations, i.e. that there are no interactions in the local predictors, which guarantees the sufficiency of pairwise interactions, i.e. the use of PMRFs. Because data was usually scarce at least with respect to the number of variables, higher-order interactions, i.e. interaction factors in the predictive models, were deliberately ignored. However, in mid-sized domains, such as in our research of depression networks incorporating descriptors, environmental factors and comorbidities, this assumption is overly restrictive.

Another open question is the remaining uncertainty at the level of model structures and in our case especially its implications for the effect strength parameters, although asymptotic consistency is proven for graphical lasso under the normality assumption [4] and sample complexity results (finite sample bounds) were reported [16], [20]. As an approximation, bootstrap methods were proposed, for its early application in the PGM field, see e.g. [21], [22]. However, the interpretation of bootstrap proportions for structural features is problematic [23]–[25]. Bayesian interpretation had already appeared in the original paper introducing lasso [26], for a detailed Bayesian interpretation, see comments from C. Holmes in [27]. Furthermore, full-fledged Bayesian Lasso and Bayesian graphical lasso were also reported in [14], [15], [28]. However, scalable, full Bayesian extension of graphical lasso for discrete data, especially with options for including higher-order interactions is currently not available.

The repertoire of R packages learning sparse graphical models using lasso contains for example the following: *glasso* [5], *parcor/adalasso* [29], *huge* [30], *qgraph* [31], *isingfit* [32], *bootnet* [33], *BDgraph* [28].

In this evaluation we will use a binary variant of graphical lasso implemented in the *bootnet* R package [33]. It uses logistic regression based predictive approximations and binary pairwise Markov Random Fields (bPRMs) [18], [19], [32].

B. Full Bayesian approaches

Full Bayesian approaches, like joint optimization of model structure and parameters, are natural candidates for the joint exploration of existential and quantitative aspects of direct dependency relations. Indeed, the joint treatment of model structures and parameters avoids the unnecessary and hard problem of model identification by performing Bayesian averaging over model structures in the exploration of effect strength relations. The full Bayesian approach and especially the idea of Bayesian averaging over model structures and the inducement of beliefs for high-level, semantic structural model properties was already present in the first proposals [34], [35]. Subsequent works suggested methods for the Markov Chain Monte Carlo based estimations of the posteriors for structural properties. First Madigan et al. proposed a Markov Chain Monte Carlo (MCMC) scheme over the space of directed acyclic graphs (DAG) using an auxiliary ordering of variables [36]. After an initial bootstrap-based approach [21], [22], Friedman and Koller developed an MCMC method over the space of variable orderings [37]. Several improvements were

suggested for the DAG-based MCMC method, such as new operators by Giudici and Castello [38] and others [39], [7]. In parallel, Koivisto and Sood introduced a method to perform exact Bayesian inference over modular features [6], [7]. A related randomized approach was presented by Pena et al. [40].

Using a full Bayesian approach based on a combination of analytic solutions for the Bayesian averaging at the parameter level and DAG- and ordering-based Bayesian averaging at the structure level, we developed the Bayesian network-based Bayesian MultiLevel Analysis (BN-BMLA) method. It uses the hierarchy of Markov Blanket Graphs (MBGs), Markov Blanket Sets (MBSs) and Memberships (MBMs) and estimates the respective a posteriori probabilities of MBGs, MBSs and MBMs. In the Bayesian statistical framework, the hierarchy of these model properties can be exploited, because their posteriors are directly induced along their topology [41]. Because the relevance of a relationship between variables implied by an effect size measure (parametric relevance) and by a dependency structure based measure (structural relevance) does not necessarily correspond to each other, it is possible to define hybrid measures that take both quantitative and qualitative aspects into account providing a more detailed view on the dependency relationships of variables. These hybrid measures were proposed using a Bayesian network formalism, as they consist of a structure representing conditional (in)dependency relationships in the form of a directed acyclic graph G , and a corresponding parametrization layer θ [42]. Such measures are the following: structure conditional Bayesian effect size measures and especially the MBG-based Bayesian effect size [42], [43], and the effect size conditional existential relevance (ECER) [44].

In this paper we will use the MBG-based Bayesian effect size measure [43].

III. APPROACH

We performed a comparative analysis of the Ising model based LASSO estimation method (IsingFit) [32] and the Bayesian model averaging based relevance analysis method [41], [43] using artificial and real data.

A. Experimental setup

Data sets consisting of 500, 1000, 5000 and 10000 complete samples were generated based on a Bayesian network model of 11 binary variables using random sampling. Figure 1 displays the structure containing 11 edges which will be referred to as the ‘reference model’ in following sections.

The aim of this comparative study was to analyze the structure learning capabilities of the involved methods. Therefore, we focused on the edge weights provided by the IsingFit R method and utilized the related bootnet package [45] to investigate their stability. Using 100 bootstrapped data sets the significance of edge weights and corresponding confidence intervals were analyzed. In case of every edge weight the proportion of data sets in which the edge weight was found significant was computed. We refer to this as *bootstrap proportion* in the subsequent evaluation.

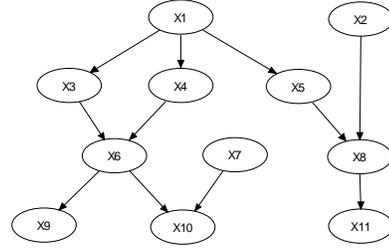


Fig. 1. The Bayesian network used to generate artificial data.

The Bayesian relevance analysis was configured to estimate edge posterior probabilities using a uniform structure prior, a Cooper-Herskovits parameter prior with 10^6 burn-in and $2 \cdot 10^6$ normal steps. In addition, Markov-blanket based Bayesian odds ratios and corresponding credible intervals were also computed [42].

Bayesian edge posteriors can be compared to bootstrap proportions as both measures indicate (although by different means) whether an edge is frequently present in possible models, i.e. its structural relevance. On the other hand, edge weights from the IsingFit method are equivalent to logistic regression coefficients and can be interpreted as effect size descriptors. Therefore, edge weights can be directly compared to other effect size measures such as the Bayesian MBG-based odds ratio.

B. Application domain

Depression became the largest single contributor to global disability [46]. It has long been recognized that depression and other medical disorders frequently co-occur in the same individual, but the elusive nature of the disease and the high level of parametric and structural uncertainties in comorbidity analyses hindered the exploration of the underlying dependency networks [33], [47]–[49]. Large scale health data sets, such as UK Biobank [50] provides a novel, unique opportunity to explore general comorbidities. Furthermore, disease-specific information available for subcohorts provides an unprecedented option to explore the dependencies of detailed descriptors of psychological and psychiatric states, together with a wide variety of variables such as personality traits, social background, lifestyle and environmental descriptors.

The investigated real-world data is a subset of a UK Biobank data set focusing on depression and related environmental, social and lifestyle factors (application No.1602). For the purposes of this study a reduced set consisting of 20 variables was selected. All variables were discretized (if needed) to produce a data set having only binary variables with a sample size of 110, 600. The IsingFit method was applied in conjunction with bootnet using 50 bootstrapped data sets. Bayesian relevance analysis was applied with similar settings as in the synthetic data analysis case. Additionally, the number of ‘parents’, i.e. incoming directed edges, was limited to 8 for each node.

C. Bayesian relevance analysis

Bayesian relevance analysis is a multivariate, model-based Bayesian method which allows the analysis of dependency relationships [41], [51]. It can be utilized as a general purpose feature subset selection method, but it also allows the refined analysis of complex dependency patterns. Bayesian relevance analysis enables the identification of relevant variables, relevant sets of variables, and interaction models of relevant variables. These are different abstraction levels of relevance which can be analyzed by evaluating the posterior probability of corresponding structural properties of an underlying Bayesian network. In order to estimate these posterior probabilities a Bayesian model-averaging framework is used.

Some properties such as Markov blankets are particularly relevant for identifying relevant variables with respect to a selected variable. The connection between Markov blankets and strong relevance was established by the theorem of Tsamardinos [52], which provides the set of conditions under which relevant structural properties are unambiguously represented. Assuming unambiguous representation, for a given structure G all the strongly relevant variables X_i with respect to Y are in the Markov blanket set of Y denoted as $MBS(Y, G)$.

A more complex structural property based on $MBS(Y, G)$ that allows to distinguish between direct relationships and interaction terms is called the *Markov blanket graph*.

Definition 1 (Markov blanket graph). “A *Markov blanket graph* $MBG(Y, G)$ of a variable Y is a subgraph of a Bayesian network structure G , which contains the nodes of the Markov blanket set of Y , that is $MBS(Y, G)$ and the incoming edges into Y and its children. Given a target node, which corresponds to the target variable Y , $MBG(Y, G)$ as a (sub)graph structure consists of nodes that are (1) parents of Y , (2) children of Y or (3) ‘other parents’ of the children of Y ” [51].

In contrast with a graph structure G which contains all the dependency relationships of variables, a Markov blanket graph $MBG(Y, G)$ includes only those relationships, in which Y is closely involved. This property makes the $MBG(Y, G)$ an ideal candidate to serve as a base for measuring effect size.

D. Bayesian statistical framework

Bayesian relevance analysis relies on a Bayesian statistical framework, more specifically on Bayesian inference over structural properties of Bayesian networks. Based on the posterior for the dependency model $P(G, \theta|D)$ related posteriors for various model properties can be derived. More specifically, the posteriors for structural properties are of interest since the main goal of this method is the discovery of the dependency structure.

In general, for an arbitrary structural property $\mathcal{F}(\mathbf{Z}, G) = f$ defined on a set of variables \mathbf{Z} the posterior probability can be defined in the following general form:

$$E_{P(G|D)}[I_{\mathcal{F}(\mathbf{Z}, G)=f}] = \sum_G P(G|D) \cdot I_{\mathcal{F}(\mathbf{Z}, G)=f}, \quad (1)$$

which means that only those $P(G|D)$ DAG structure posteriors are considered for which the indicator function $I_{\mathcal{F}(\mathbf{Z}, G)=f} = 1$, i.e. the structural property $\mathcal{F}(\mathbf{Z}, G) = f$ holds [41]. In other words, the general task of Bayesian relevance analysis would be to compute the expectation function of structural properties such as Markov blanket graphs over the space of DAG structures. However, the exact computation of such posteriors is generally infeasible due to the high cardinality of possible DAG structures.

In order to efficiently estimate posteriors of structural properties of Bayesian networks, we proposed specialized MCMC methods [51], [53]. The latter is a DAG MCMC method which can be applied for the estimation of the posterior of any structural property $\mathcal{F}(U, G) = f$ provided that an efficient computation exists for the evaluation of the property and for the unnormalized posterior $\tilde{P}(G|D)$ [35], [54].

The unnormalized posterior is utilized by an MCMC sampling method which performs a random walk in the space of DAGs. The transition between states (i.e. between various DAG structures) is facilitated by DAG operators which delete, insert or invert edges in the DAG structure [38]. The default strategy is to apply these operators with a uniform probability in the proposal distribution.

In a standard scenario, the MCMC process reaches a stationary state, i.e. convergence is established after a number of steps. This interval is called the burn-in, during which the sampled states of the random walk are discarded. Only the samples collected after the burn-in are utilized for the computation of posteriors.

The MCMC process results in a sequence of DAG structures which is used to estimate the posteriors of structural properties. The crucial step for the relevance analysis is the identification of the Markov blanket graph for each target variable given each DAG structure visited during the random walk. Note that a joint Markov blanket graph of multiple targets can also be identified at each MCMC step. Apart from Markov blanket graphs, other structural properties can also be identified and evaluated based on the currently encountered DAG structure. After the identification and evaluation, statistics related to the investigated properties are updated.

Following the MCMC process an additional computational phase is executed. This step consist of computing the relative frequency of each instantiation of the investigated structural properties $\mathcal{F}(U, G) = f$ which leads to the estimate of the posterior probability values given M samples:

$$E_{P(G|D)}[I_{\mathcal{F}(\mathbf{Z}, G)=f}] \approx \frac{1}{M} \sum_{i=1}^M I_{\mathcal{F}(\mathbf{z}, G_i)=f}, \quad (2)$$

E. A structure conditional Bayesian effect size measure

Bayesian networks as a model class allows the analysis of dependency relationships of variables from both qualitative (dependency structure based) and quantitative (effect size based) aspects. Investigating the structural properties of the underlying DAG structure enables qualitative assessment, while computing effect size measures based on the parametric layer

allows quantitative assessment of relationships. In addition, it is possible to define such measures that take both aspects into consideration.

The key notion is that if a variable X_i is a member of the Markov blanket set of the target variable Y , then the probability of a specific value of the target Y can be estimated based on the $MBG(Y, G)$ and a specific instantiation of X_i [55]. Therefore, instead of taking into account the whole structure G , the conditional probabilities required for odds ratio computation can be estimated using only $MBG(Y, G)$.

In the frequentist framework an odds is defined as the ratio of conditional probabilities [56].

Definition 2 (Odds). *Let X_1, X_2, \dots, X_n denote discrete variables that have $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ states respectively, and let Y denote the target variable with y_1, \dots, y_q possible states. Then $X_i = x_{i_1}$ denotes variable X_i in state x_{i_1} , and an odds is defined as*

$$\text{Odds}(X_i^{(j)}, Y^{(m,n)}) = \frac{p(Y = y_m | X_i = x_{i_j})}{p(Y = y_n | X_i = x_{i_j})}. \quad (3)$$

This means that the odds is computed for a given value of X_i (denoted as $X_i^{(j)}$) using a selected pair of values of Y (denoted as $Y^{(m,n)}$). Consequently, the odds can be used to compute the odds ratio for a pair of values of variable X_i (denoted as $X_i^{(k,j)}$) given that both odds were computed for the same pair of values of Y (denoted as $Y^{(m,n)}$) [56].

Definition 3 (Odds ratio). *An odds-ratio for variable X_i between states x_{i_j} and x_{i_k} is given as*

$$\text{OR}(X_i^{(k,j)}, Y^{(m,n)}) = \frac{\text{Odds}(X_i^{(k)}, Y^{(m,n)})}{\text{Odds}(X_i^{(j)}, Y^{(m,n)})} \quad (4)$$

For example, in case of a case-control study the target variable Y is a binary disease indicator such that $Y = 0$: non-affected (control), $Y = 1$: affected (case), and discrete variables X_1, X_2, \dots, X_n represent investigated clinical, social or environmental factors. Then $\text{OR}(X_i^{(1,0)}, Y^{(1,0)})$ denotes an odds ratio of variable X_i with respect to target Y , which indicates how $X_i = 1$ influences disease susceptibility compared to $X_i = 0$. In other words, it quantifies the effect of variable X_i on Y .

In the Bayesian framework, an odds ratio is a random variable with a corresponding distribution depending on the parametrization layer learned from the data. Accordingly, the structure conditional odds ratio denoted as $\text{OR}(X_i, Y | \theta, G)$ is a random variable with a distribution defined by $p(\theta, G | D)$, in which parametrization θ depends on a specific structure G , and D denotes data. As mentioned earlier, a more practical approach is to utilize Markov blanket graphs instead of the whole G which leads to the Markov blanket graph based Bayesian odds ratio [42], [43]:

Definition 4 (Markov blanket graph based Bayesian odds ratio). *The MBG-based Bayesian odds ratio (MBG-BOR) is computed by averaging over the estimates of odds ratios based on possible MBGs as follows*

$$\text{MBG-BOR}(X_i, Y | D) = \sum_{j=1}^n \text{OR}(X_i, Y | \text{MBG}_j(Y, G)) \cdot p(\text{MBG}_j(Y, G) | D) \cdot I_{(X_i \in \text{MBG}_j(Y, G))}, \quad (5)$$

where m is the number of MBGs with a posterior $p(\text{MBG}_j(Y, G) | D) > 0$. The indicator function $I_{(X_i \in \text{MBG}_j(Y, G))}$ is 1 if $X_i \in \text{MBG}_j(Y, G)$ and 0 otherwise.

IV. RESULTS

First, we present the evaluation of the Ising model based LASSO estimation method (IsingFit) [32] and the Bayesian relevance analysis method [41], [43] on artificial data, then we describe results corresponding to the real data.

One of the goals of this study is the comparison of bootstrap proportions of structural properties (e.g. edges) against corresponding posteriors. Note that the usual visualization of pairwise Markov Random Fields shows the pairwise partial correlations (β coefficients) by edge coloring and statistical significances by edge width, whereas Bayesian network based dependency maps are generally visualized by showing only edge posteriors as quantifications of structural (existential) uncertainties. Specifically, in graphical lasso the averages of β_{i-j} and β_{j-i} , are shown between node i and j , whereas in a Bayesian approach edge widths represent the sum of edge posteriors $p(i \rightarrow j | D)$ and $p(j \rightarrow i | D)$.

Here however, we wish to investigate both quantitative (parametric) and qualitative (structural) aspects of both methods, thus for the sake of comparability we defer from these conventions in certain cases, and indicate the correspondence between graph features and represented measures.

A. Evaluation based on artificial data

Performance measures are presented in Table I for all artificial data sets, and evaluated according to the reference model. SEN , SPC , denote sensitivity and specificity respectively. In addition, the results of a chi-square based association test (using a $\alpha = 0.05$ significance threshold) was also included as a reference. In case of IsingFit and edge was termed as identified if the edge weight was found significant based on its confidence interval. In case of a Bayesian method a decision function can be defined, however there is no strict threshold of relevance for posterior probabilities. For the purposes of this evaluation we chose the arbitrary threshold of $pr > 0.5$. Thus an edge was termed as identified if it had a larger posterior probability.

Results indicate that 500 samples were insufficient to learn the structure of the reference model properly for both the IsingFit and the Bayesian relevance analysis methods, leading to low sensitivity scores particularly in case of the former. In case of 1000 samples the Bayesian method found all the edges of the reference model and reduced the number of false edges to zero. The performance of the IsingFit method improved, but still missed a few edges. Figure 2 displays the identified edges for both methods given the 1000 samples. The first structure corresponds to the results of the Bayesian method, where the

TABLE I
COMPARISON OF THE PERFORMANCE OF LASSO-BASED AND BAYESIAN METHODS BASED ON THE ARTIFICIAL DATA.

| Sample Size | Association | | LASSO - IsingFit | | Bayesian | |
|-------------|-------------|------|------------------|------|----------|------|
| | SEN | SPC | SEN | SPC | SEN | SPC |
| 500 | 0.82 | 0.93 | 0.45 | 1.00 | 0.73 | 0.95 |
| 1000 | 1.00 | 0.91 | 0.55 | 1.00 | 1.00 | 1.00 |
| 5000 | 1.00 | 0.80 | 0.82 | 0.98 | 1.00 | 1.00 |
| 10000 | 1.00 | 0.77 | 1.00 | 0.98 | 1.00 | 1.00 |

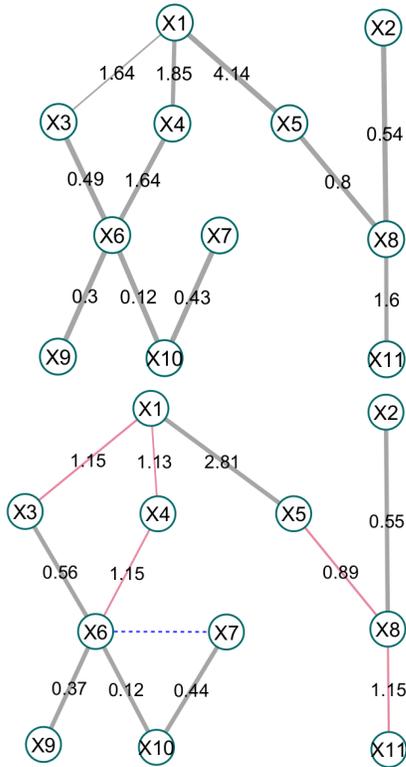


Fig. 2. Evaluation of methods based on artificial data

edge width is proportional to the posterior probability, and edge labels represent MBG-based Bayesian odds ratios. Note that the directionality of edges is omitted.

The second structure corresponds to the IsingFit method, where red colored edges indicate missing edges that were found non-significant. Edge labels represent (the exponential of) edge weights. As data size increased the sensitivity of IsingFit gradually improved, however its specificity decreased. This is due to the fact that the V-structure [1] formed by variables $X_6 \rightarrow X_{10} \leftarrow X_7$ represents a multivariate interaction that can only be represented with a triangle in a Markov random field thus requiring an additional edge between $X_6 - X_7$ (marked blue in Figure 2).

In summary, Bayesian relevance analysis performs better in the current scenario than IsingFit, which is partially due to the fact that the data sets were generated from a Bayesian network.

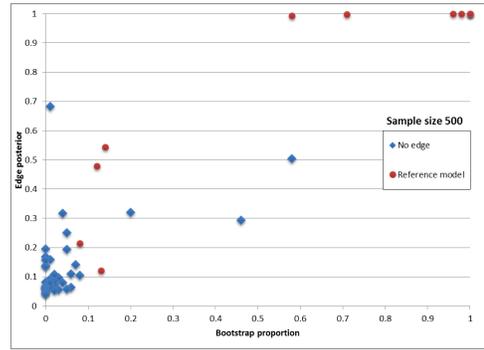


Fig. 3. Comparison of edge posteriors and bootstrap proportions given the artificial data set of 500 samples.

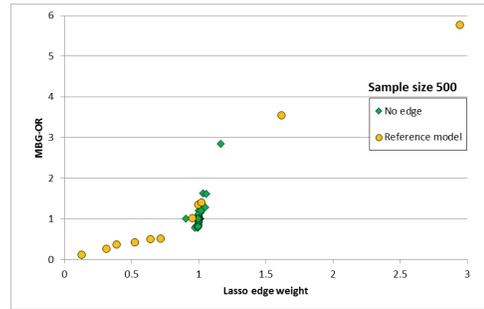


Fig. 4. Comparison of IsingFit edge weights and MBG-based Bayesian odds ratios given the artificial data set of 500 samples.

Note that the association test had a high sensitivity, but its specificity gradually decreased due to the increasing number of false positive edges. Correction methods can alleviate such symptoms of multiple testing to a degree, but they may also impact sensitivity [57].

Furthermore, we performed additional analyses to compare the properties of the investigated methods. Figure 3 presents the comparison of bootstrap proportions and edge posteriors for the data set with 500 samples.

These results indicate that among reference model edges (marked as red dots) the IsingFit method used in conjunction with bootnet retained a high uncertainty with respect to several of these edges, which is reflected by considerably low bootstrap ratios. Corresponding edge posteriors are higher in some of the cases, however the number of false positives (i.e. edges that are not present in the reference model) is also higher in case of the Bayesian relevance analysis method.

Regarding the parametric aspect of variable relationships we compared edge weights of the IsingFit method and the MBG-based Bayesian odds ratios shown in Figure 4. Results indicate that both methods assign approximately neutral effect sizes to most non-relevant edges (marked as green diamonds). In case of relevant edges (marked as yellow dots), the Bayesian odds ratio tends to assign a higher score (in case of $OR > 1$ and lower score for $OR < 1$) for a relationship between structurally relevant variables. This behavior is in accordance with its definition, i.e. utilizing structural aspects of relevance.

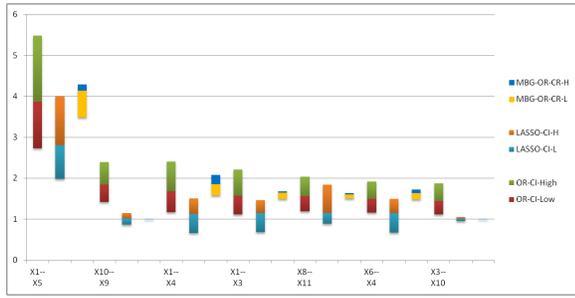


Fig. 5. Comparison of confidence and credible intervals of effect size measures for a selected set of variable relationships. Lasso-CI-H and CI-L denote the higher and lower segments of the confidence interval of edge weights of the InsingFit method. Similarly OR-CI-H and OR-CI-L correspond to the confidence interval of the frequentist odds ratio, while MBG-OR-CR-H and MBG-OR-CR-L denote the credible interval of Bayesian odds ratio. The junction of the higher and lower segments is at the mean value of the corresponding odds ratio.

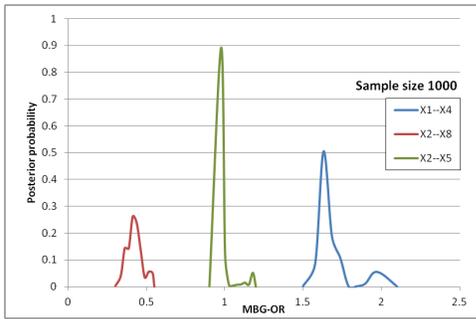


Fig. 6. An example of MBG-based Bayesian odds ratio distributions. In case of variables $X_1 - X_4$ and $X_2 - X_8$ the Bayesian odds ratio is relevant, the corresponding credible intervals are compact and do not intersect with the neutral effect size of 1. In contrast, the credible interval of $X_2 - X_5$ contains the neutral effect size and thus it is not relevant.

Furthermore, we compared the confidence intervals of edge weights and credible intervals of MBG-based Bayesian odds ratios shown in Figure 5. As a reference frequentist odds ratios and corresponding confidence intervals are also presented. Results indicate that InsingFit edge weights generally underestimate frequentist odds ratios and have a less wide confidence interval. In contrast, the mean value of Bayesian odds ratios is close to the corresponding frequentist odds ratios in case of structurally relevant variable relationships and have a compact credible interval. On the other hand, in structurally irrelevant cases (e.g. $X_3 - X_{10}$, $X_9 - X_{10}$) the mean value is reduced to the neutral effect size and the credible interval collapses to a minimal size.

An additional feature of MBG-based Bayesian odds ratios is that their distribution can be analyzed allowing further insights. An example is presented in Figure 6. The Markov blanket graph, which serves as its basis, encodes multivariate dependencies which can manifest multi-modal distributions. This can indicate that the effect size of the investigated variable relationship depends on a given context. Thus the analysis of such Bayesian odds ratio distributions may direct further investigations.

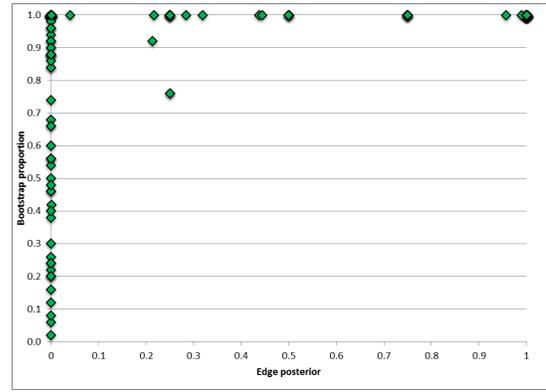


Fig. 7. Scatter plot of Bayesian edge posteriors and bootstrap proportions related to edge weights of the InsingFit - bootnet method.

B. Comparison of results based on real data

The main purpose of this real data based comparison was to observe the characteristics of both the InsingFit and Bayesian relevance analysis methods in an association rich domain. Furthermore, multivariate relationships are key features of such domains. Since InsingFit is a logistic regression like approximation excluding interactions, while Bayesian relevance analysis is a model-based multivariate method taking (even) higher-order interactions into account, considerable differences in results are likely to appear.

Figure 7 compares bootstrap proportions and edge posteriors related to the UKB data set. The remarkable feature of this scatter plot is that there appears to be several cases in which Bayesian edge posteriors are zero, whereas bootstrap proportions are maximal (i.e. those variable relationships were detected in all bootstrapped data sets).

In addition, Figure 8 presents the same edge posteriors and bootstrap proportions in a pairwise way for all possible edges. These results indicate that there are almost 40 edges having zero posterior probability and maximal bootstrap proportion. A possible explanation of this phenomenon is that some of these edges are ‘required’ to represent the V-structure like relationships in a Markov random field, similarly to the case explained at the comparative study based on artificial data. Another possible explanation is that there can be such dependency relationships that Bayesian networks cannot adequately represent. Either way this phenomenon requires further investigation.

Figure 9 displays InsingFit edge weights and MBG-based Bayesian odds ratios as log odds ratios. The effect size assigned to particular edges are similar in the majority of the cases, although the Bayesian odds ratios are moderately higher (or lower if $OR < 1$). An exception to this phenomenon is the set of edges (depicted as blue markers) which have a neutral Bayesian odds ratio ($\log OR = 0$) and a moderate edge weight. These correspond to the previously mentioned edges with zero posterior probability and high bootstrap proportion.

Finally, Figure 10 presents variable relationships learned by both InsingFit and Bayesian relevance analysis methods. The

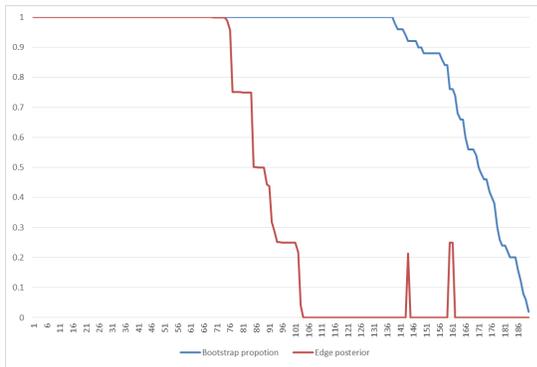


Fig. 8. Comparison of Bayesian edge posteriors and bootstrap proportions related to edge weights of the IsingFit - bootnet method.

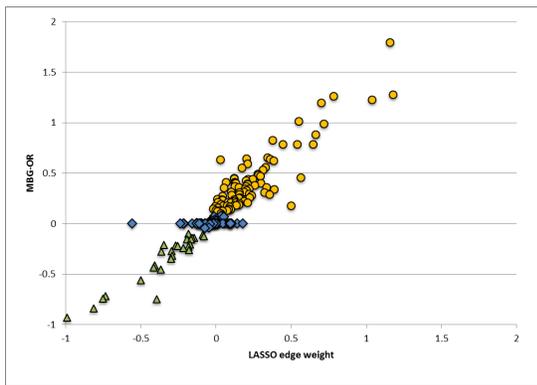


Fig. 9. Comparison of Bayesian and Lasso based approach. Lasso edge weights and MBG-based Bayesian odds ratios.

first structure displays results of the Bayesian method, where edge color refers to edge posteriors (yellow: $0.5 < pr < 0.75$, green: $0.75 < pr < 0.95$, black: $pr < 0.95$), and edge width corresponds to MBG-based Bayesian odds ratios. The second structure is the result of IsingFit, where edge width corresponds to edge weights. Edges having bootstrap proportion lower than 1 were omitted. Some of the relationships having high effect size can be identified in both structures such as the *Body fat - Obesity - Metabolic rate* relationships, and the *Satisfaction- Neuroticism - Current depression* relations. In general, IsingFit learned a larger number of edges compared to Bayesian relevance analysis, which might be due to the different representational capabilities. The top 20 variable relationships having the highest effect size are displayed in Table II. The table also lists the rank of each relationship according to the investigated measures. There is minimal variation in the ranking order of the top 10 results. Among higher ranks, however, there are considerable differences.

Both methods have additional measures that enable a refined analysis of such models. Bayesian relevance analysis methodology provides detail oriented analysis tools that allow the investigation of selected relationship sets and relationship types.

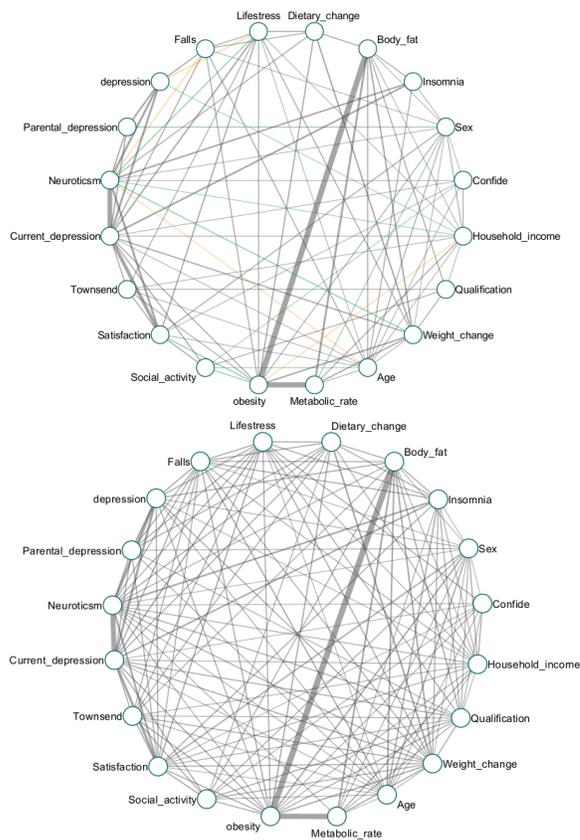


Fig. 10. Variable relationships learned from real data by Bayesian relevance analysis (upper) and IsingFit (lower) methods.

TABLE II
TOP 20 VARIABLE RELATIONSHIPS BASED ON THE UKB DATA SET HAVING THE HIGHEST EFFECT SIZE ACCORDING TO ISINGFIT AND BAYESIAN RELEVANCE ANALYSIS METHODS.

| Relationship | MBG-OR | Edge weight | Rank - MBG-OR | Rank - Edge w. |
|--|--------|-------------|---------------|----------------|
| Obesity – Body fat | 42.70 | 32.60 | 1 | 1 |
| Metabolic rate – Obesity | 27.96 | 16.63 | 2 | 2 |
| Neuroticism – Current depression | 11.32 | 7.03 | 3 | 3 |
| Satisfaction – Current depression | 6.02 | 3.18 | 4 | 5 |
| Neuroticism – Depression(lifetime) | 3.59 | 3.24 | 5 | 4 |
| Satisfaction – Neuroticism | 3.53 | 2.19 | 6 | 7 |
| Current depression – Depression(lifetime) | 3.41 | 2.82 | 7 | 6 |
| Metabolic rate – Body fat | 3.31 | 2.01 | 8 | 9 |
| Current depression – Insomnia | 2.76 | 1.73 | 9 | 13 |
| Parental depression – Depression(lifetime) | 2.69 | 2.05 | 10 | 8 |
| Neuroticism – Insomnia | 2.42 | 1.94 | 11 | 10 |
| Weight change – Obesity | 2.29 | 1.46 | 12 | 19 |
| Dietary change – Weight change | 2.20 | 1.91 | 13 | 11 |
| Weight change – Body fat | 2.20 | 1.71 | 14 | 14 |
| Life stress – Current depression | 2.19 | 1.56 | 15 | 16 |
| Metabolic rate – Weight change | 1.93 | 1.41 | 16 | 22 |
| Parental depression – Current depression | 1.91 | 1.22 | 17 | 42 |
| Satisfaction – Life stress | 1.90 | 1.43 | 18 | 20 |
| Parental depression – Neuroticism | 1.86 | 1.46 | 20 | 18 |
| Current depression – Weight change | 1.81 | 1.23 | 21 | 40 |
| Age – Body fat | 1.58 | 1.76 | 28 | 12 |
| Satisfaction – Sex | 1.41 | 1.47 | 49 | 17 |
| Age – Sex | 1.34 | 1.43 | 57 | 21 |
| Obesity – Sex | 1.20 | 1.65 | 83 | 15 |

V. CONCLUSION

The lack of general guides for the MCMC methods, the high computational complexity and additional uncertainty measures of MCMC convergence and MCMC confidence estimates prompted widespread interest in the application of graphical

lasso, because of its robust statistical properties, sparsity and low computational complexity, see e.g. [33], [48], [49].

As the reported earlier applications showed, the hybrid, full Bayesian approach is scalable to mid-sized problems with hundreds of variables [58], [59]. Although the computational complexity is significantly higher for hybrid, full Bayesian inference (which applies MCMC methods at the structure level and analytic integration at the parameter level), than for a graphical lasso, even within a standard bootstrap framework. In such mid-sized problems the advantages of the full Bayesian inference using Bayesian networks (BNs) versus the usage of graphical lasso using pairwise Markov Random Fields, are threefold:

- 1) The hybrid, full Bayesian approach relying on Bayesian networks with multinomial local models for complete discrete data, performs Rao-Blackwellization using analytic integration at the parameter level.
- 2) The full Bayesian approach supports non-linear dependencies including higher-order interactions (cf. the assumptions of linearity and no interactions in graphical LASSO and its logistic regression based approximations). As we showed in this paper, the underlying representation of pairwise MRFs in graphical LASSO can suffer from spurious dependencies resulting from such high-order interactions.
- 3) Bayesian model averaging (BMA) allows a principled solution to cope with a high level of uncertainty at the structural level. For example, if the posterior distribution of possible model structures is ‘flat’ and there are no dominant maximum a posteriori models, then Bayesian model averaging usually provides estimates with less certainty at parameter levels. Additionally, through the posterior over structures, there is a consistent Bayesian semantics for uncertainties over structural properties. Note that the computationally less complex bootstrap approaches have rather indirect interpretations for discrete variables representing structural properties, e.g. in phylogenetic tree inference.

Note that the Bayesian approach performs model averaging, i.e. it averages over (dominantly sparse) models. Consequently, the resulting edge posterior map is not sparse in a strict sense as it contains vanishing posteriors as well. However, in a decision theoretic situation (e.g. reporting with a given cost function) an appropriate threshold can be selected normatively to achieve minimum expected loss. This means that contrary to the lasso approach, where complexity regularization is an integral part of model fitting, the level of sparsity in the Bayesian approach can be quantitatively controlled off-line after the analysis, which is crucial for sharing uncertain knowledge.

ACKNOWLEDGMENT

This research has been conducted using the UK Biobank Resource under Application Number 1602. This study has been supported by the National Development Agency (KTIA_NAP_13-1-2013-0001), Hungarian Brain Research

Program - Grant No. KTIA_13_NAP-A-II/14; the MTA-SE Neuropsychopharmacology and Neurochemistry Research Group, Hungarian Academy of Sciences, Semmelweis University; the Hungarian Academy of Sciences and the Hungarian Brain Research Program - Grant No. KTIA_NAP_13-2-2015-0001 (MTA-SE-NAP B Genetic Brain Imaging Migraine Research Group); OTKA 119866 and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (P. Antal), and by the Manchester Academic Health Science Centre. The sponsors funded the work, but had no further role in the design of the study, in data collection or analysis, in the decision to publish, or in the preparation, review, or approval of the manuscript.

REFERENCES

- [1] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [2] D. J. Spiegelhalter, A. Dawid, S. Lauritzen, and R. Cowell, “Bayesian analysis in expert systems,” *Statistical Science*, vol. 8, no. 3, pp. 219–283, 1993.
- [3] C. Glymour, R. Scheines, and P. Spirtes, “Exploring causal structure with the tetrad program,” *Sociological methodology*, vol. 18, pp. 411–448, 1988.
- [4] N. Meinshausen and P. Bühlmann, “High-dimensional graphs and variable selection with the lasso,” *The annals of statistics*, pp. 1436–1462, 2006.
- [5] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [6] M. Koivisto and K. Sood, “Exact bayesian structure discovery in bayesian networks,” *Journal of Machine Learning Research*, vol. 5, pp. 549–573, 2004.
- [7] T. Niinimäki, P. Parviainen, and M. Koivisto, “Partial order mcmc for structure discovery in bayesian networks,” in *Proc. of the Twenty-Seventh Conf. on Uncertainty in Artificial Intelligence (UAI-11), Barcelona, Spain, July 14-17, 2011*, 2011, pp. 557–564.
- [8] K. Y. Yeung, C. Fraley, W. C. Young, R. Bumgarner, and A. E. Raftery, “Bayesian model averaging methods and r package for gene network construction,” in *Big data analytic technology for Bioinformatics and health informatics (KDDBHI), workshop at the 20th ACM SIGKDD conference on knowledge discovery and data mining (KDD), New York, 2014*.
- [9] A. P. Dempster, “Covariance selection,” *Biometrics*, pp. 157–175, 1972.
- [10] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.
- [11] S. L. Lauritzen, *Graphical Models*. UK, Clarendon: Oxford, 1996.
- [12] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [13] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [14] T. Park and G. Casella, “The bayesian lasso,” *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.
- [15] H. Wang *et al.*, “Bayesian graphical lasso models and efficient posterior computation,” *Bayesian Analysis*, vol. 7, no. 4, pp. 867–886, 2012.
- [16] P.-L. Loh, M. J. Wainwright *et al.*, “Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses.” in *NIPS*, 2012, pp. 2096–2104.
- [17] B. Fellinghauer, P. Bühlmann, M. Ryffel, M. Von Rhein, and J. D. Reinhardt, “Stable graphical model estimation with random forests for discrete, continuous, and mixed variables,” *Computational Statistics & Data Analysis*, vol. 64, pp. 132–152, 2013.
- [18] R. Strobl, G. Stucki, E. Grill, M. Müller, and U. Mansmann, “Graphical models illustrated complex associations between variables describing human functioning,” *Journal of clinical epidemiology*, vol. 62, no. 9, pp. 922–933, 2009.
- [19] R. Strobl, E. Grill, and U. Mansmann, “Graphical modeling of binary data using the lasso: a simulation study,” *BMC medical research methodology*, vol. 12, no. 1, p. 16, 2012.

- [20] N. Friedman and Z. Yakhini, "On the sample complexity of learning bayesian networks," in *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1996, pp. 274–282.
- [21] N. Friedman, M. Goldszmidt, and A. Wyner, "Data analysis with bayesian networks: A bootstrap approach," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 196–205.
- [22] N. Friedman, M. Goldszmidt, and A. J. Wyner, "On the application of the bootstrap for computing confidence measures on features of induced bayesian networks," in *AISTATS*, 1999.
- [23] J. Felsenstein and H. Kishino, "Is there something wrong with the bootstrap on phylogenies? a reply to hillis and bull," *Systematic Biology*, vol. 42, no. 2, pp. 193–200, 1993.
- [24] M. E. Alfaro, S. Zoller, and F. Lutzon, "Bayes or bootstrap? a simulation study comparin the performance of bayesian mcmc sampling and bootstrapping in assesing phylogenetic confidence," *Mol. Biol. Evol.*, vol. 20, no. 2, pp. 255–266, 2003.
- [25] C. J. Douady, F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. Douzery, "Comparison of bayesian and maximum likelihood bootstrap measures of phylogenetic reliability," *Molecular Biology and Evolution*, vol. 20, no. 2, pp. 248–254, 2003.
- [26] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [27] —, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, 2011.
- [28] A. Mohammadi, E. C. Wit *et al.*, "Bayesian structure learning in sparse gaussian graphical models," *Bayesian Analysis*, vol. 10, no. 1, pp. 109–138, 2015.
- [29] N. Krämer, J. Schäfer, and A.-L. Boulesteix, "Regularized estimation of large-scale gene association networks using graphical gaussian models," *BMC bioinformatics*, vol. 10, no. 1, p. 384, 2009.
- [30] T. Zhao, H. Liu, K. Roeder, J. Lafferty, and L. Wasserman, "The huge package for high-dimensional undirected graph estimation in r," *Journal of Machine Learning Research*, vol. 13, no. Apr, pp. 1059–1062, 2012.
- [31] S. Epskamp, A. O. Cramer, L. J. Waldorp, V. D. Schmittmann, D. Borsboom *et al.*, "qgraph: Network visualizations of relationships in psychometric data," *Journal of Statistical Software*, vol. 48, no. 4, pp. 1–18, 2012.
- [32] C. D. van Borkulo, D. Borsboom, S. Epskamp, T. F. Blanken, L. Boschloo, R. A. Schoevers, and L. J. Waldorp, "A new method for constructing networks from binary data," *Scientific reports*, vol. 4, p. 5918, 2014.
- [33] E. I. Fried, S. Epskamp, R. M. Nesse, F. Tuerlinckx, and D. Borsboom, "What are 'good' depression symptoms? comparing the centrality of dsm and non-dsm symptoms of depression in a network analysis," *Journal of affective disorders*, vol. 189, pp. 314–320, 2016.
- [34] W. L. Buntine, "Theory refinement of Bayesian networks," in *Proc. of the 7th Conf. on Uncertainty in Artificial Intelligence (UAI-1991)*. Morgan Kaufmann, 1991, pp. 52–60.
- [35] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309–347, 1992.
- [36] D. Madigan, S. A. Andersson, M. Perlman, and C. T. Volinsky, "Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs," *Comm.Statist. Theory Methods*, vol. 25, pp. 2493–2520, 1996.
- [37] N. Friedman and D. Koller, "Being Bayesian about network structure," *Machine Learning*, vol. 50, pp. 95–125, 2003.
- [38] P. Giudici and R. Castelo, "Improving Markov Chain Monte Carlo model search for data mining," *Machine Learning*, vol. 50, pp. 127–158, 2003.
- [39] M. Grzegorzcyk and D. Husmeier, "Improving the structure mcmc sampler for bayesian networks by introducing a new edge reversal move," *Machine Learning*, vol. 71, no. 2-3, pp. 265–305, 2008.
- [40] J. Pena, R. Nilsson, J. Björkregren, and J. Tegnér, "Towards scalable and data efficient learning of markov boundaries," *International Journal of Approximate Reasoning*, vol. 45, pp. 211–232, 2007.
- [41] P. Antal, A. Millinghoffer, G. Hullám, C. Szalai, and A. Falus, "A bayesian view of challenges in feature selection: Feature aggregation, multiple targets, redundancy and interaction," *JMLR Proceeding*, vol. 4, pp. 74–89, 2008.
- [42] G. Hullám and P. Antal, "The effect of parameter priors on bayesian relevance and effect size measures," *Periodica Polytechnica. Electrical Engineering and Computer Science*, vol. 57, no. 2, p. 35, 2013.
- [43] G. Hullam and P. Antal, "Estimation of effect size posterior using model averaging over bayesian network structures and parameters," in *Proceedings of the 6th European Workshop on Probabilistic Graphical Models. Granada, Spain*, 2012, pp. 147–154.
- [44] —, "Towards a bayesian decision theoretic analysis of contextual effect modifiers," in *European Workshop on Probabilistic Graphical Models*. Springer, 2014, pp. 222–237.
- [45] S. Epskamp, D. Borsboom, and E. I. Fried, "Estimating psychological networks and their accuracy: a tutorial paper," *arXiv preprint arXiv:1604.08462*, 2016.
- [46] W. H. Organization *et al.*, "Depression and other common mental disorders: global health estimates," 2017.
- [47] A. O. Cramer, L. J. Waldorp, H. L. van der Maas, and D. Borsboom, "Complex realities require complex theories: Refining and extending the network approach to mental disorders," *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 178–193, 2010.
- [48] C. van Borkulo, L. Boschloo, D. Borsboom, B. W. Penninx, L. J. Waldorp, and R. A. Schoevers, "Association of symptom network structure with the course of depression," *JAMA psychiatry*, vol. 72, no. 12, 2015.
- [49] C. Beard, A. Millner, M. Forgeard, E. Fried, K. Hsu, M. Treadway, C. Leonard, S. Kertz, and T. Björgvinsson, "Network analysis of depression and anxiety symptom relationships in a psychiatric sample," *Psychological medicine*, vol. 46, no. 16, p. 3359, 2016.
- [50] N. Allen, C. Sudlow, P. Downey, T. Peakman, J. Danesh, P. Elliott, J. Gallacher, J. Green, P. Matthews, J. Pell *et al.*, "Uk biobank: Current status and what it means for epidemiology," *Health Policy and Technology*, vol. 1, no. 3, pp. 123–126, 2012.
- [51] P. Antal, G. Hullám, A. Gézsi, and A. Millinghoffer, "Learning complex bayesian network features for classification," in *Proc. of Third European Workshop on Probabilistic Graphical Models*, 2006, pp. 9–16.
- [52] I. Tsamardinos and C. Aliferis, "Towards principled feature selection: Relevancy, filters, and wrappers," in *Proc. of the Ninth International Workshop on Artificial Intelligence and Statistics*. Key West, Florida, USA: Morgan Kaufmann Publishers, 2003, pp. 334–342.
- [53] P. Antal, A. Millinghoffer, G. Hullám, C. Szalai, and A. Falus, "A bayesian multilevel analysis of feature relevance," in *Workshop on New challenges for feature selection in data mining and knowledge discovery (FSDM 2008) at The 19th European Conference on Machine Learning (ECML 2008)*, 2008, pp. 74–89.
- [54] D. Heckerman and D. Geiger, "Likelihoods and parameter priors for Bayesian networks," 1995, tech. Rep. MSR-TR-95-54, MicroSoft Research.
- [55] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann, 1988.
- [56] D. Sheskin, "Parametric and nonparametric statistical procedures," *Boca Raton: CRC*, 2000.
- [57] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc.*, vol. 57, no. 1, pp. 289–300, 1995.
- [58] I. Ungvari, G. Hullam, P. Antal, P. Kiszel, A. Gezsi, and C. S. *et al.*, "Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance," *PLoS ONE*, vol. 7, no. 2, p. e33573, 2012.
- [59] O. Lautner-Csorba, A. Gézsi, D. J. Erdélyi, G. Hullám, P. Antal, Á. F. Semsei, N. Kutszegi, G. Kovács, A. Falus, and C. Szalai, "Roles of genetic polymorphisms in the folate pathway in childhood acute lymphoblastic leukemia evaluated by bayesian relevance and effect size analysis," *PloS one*, vol. 8, no. 8, p. e69843, 2013.