

Psychological Methods

A Tutorial on Bayesian Networks for Psychopathology Researchers

Giovanni Briganti, Marco Scutari, and Richard J. McNally

Online First Publication, February 3, 2022. <http://dx.doi.org/10.1037/met0000479>

CITATION

Briganti, G., Scutari, M., & McNally, R. J. (2022, February 3). A Tutorial on Bayesian Networks for Psychopathology Researchers. *Psychological Methods*. Advance online publication. <http://dx.doi.org/10.1037/met0000479>

A Tutorial on Bayesian Networks for Psychopathology Researchers

Giovanni Briganti^{1, 2}, Marco Scutari³, and Richard J. McNally¹

¹ Department of Psychology, Harvard University

² Laboratoire de Psychologie Médicale et Addictologie, Université libre de Bruxelles

³ Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland

Abstract

Bayesian Networks are probabilistic graphical models that represent conditional independence relationships among variables as a directed acyclic graph (DAG), where edges can be interpreted as causal effects connecting one causal symptom to an effect symptom. These models can help overcome one of the key limitations of partial correlation networks whose edges are undirected. This tutorial aims to introduce Bayesian Networks to identify admissible causal relationships in cross-sectional data, as well as how to estimate these models in R through three algorithm families with an empirical example data set of depressive symptoms. In addition, we discuss common problems and questions related to Bayesian networks. We recommend Bayesian networks be investigated to gain causal insight in psychological data.

Translational Abstract

In the last decade, the network framework for the study of mental disorders has emerged as a new way of investigating mental disorders as issuing from interactions among their constituent symptoms. Network analysis is the statistical aspect of this framework, as researchers use nodes (symptoms) and edges (connections between symptoms) to model disorders: Usually, network structures encode pairwise interactions among symptoms. In this study, we introduce Bayesian networks, models that can identify admissible causal relationships in cross-sectional data, as well as a tutorial for applied researchers on how to estimate those models in R. In addition, we discuss common problems and questions related to Bayesian network models.


Keywords: Bayesian networks, directed acyclic graphs, network modeling, causal inference, tutorial

The last decade has seen the emergence of a new framework for the study of psychopathology: network theory, which conceptualizes an episode of disorder as issuing from interactions among its constituent symptoms (Borsboom, 2017; Borsboom & Cramer, 2013). Network theory comes with a set of statistical techniques called *network analysis*, where the *networks* that represent interactions among symptoms are composed of *nodes* (the symptoms themselves) and *edges* (connections among symptoms). In contrast to social networks, whose nodes

(persons) and edges (e.g., friendships) are observable, psychopathology networks require statistical estimation to discern the unobservable connections between symptoms. Most network structures in the psychopathology field encode pairwise interactions among symptoms (Epskamp & Fried, 2018), and use a general class of models called pairwise Markov random fields.

Gaussian graphical models (GGMs), also called partial correlation networks, are the most common way of estimating network structures in many areas of psychopathology, such as posttraumatic stress disorder (Fried et al., 2018; McNally et al., 2015); depression (Briganti et al., 2021; Fried et al., 2016) and bipolar disorders (Curtiss et al., 2019). Other pairwise Markov random fields include Ising models to estimate pairwise interactions from binary data (Briganti & Linkowski, 2020; Ising, 1925; Kruis & Maris, 2016; van Borkulo et al., 2014) and mixed graphical models to deal with mixed data (Haslbeck & Fried, 2017; Haslbeck & Waldorp, 2016). However, the causal interpretation of these models is limited: because their edges are *undirected*, it is impossible to tell whether symptom X is more likely to *cause* or *be caused* by symptom Y because edges have no direction and thus cannot encode this information. Furthermore, assuming a partial correlation network when the underlying model contains directed edges can lead to spurious causal connections.

We can overcome these problems with the help of Bayesian networks (BNs), which are derived from the principles of causal

Giovanni Briganti  <https://orcid.org/0000-0002-4038-3363>

Giovanni Briganti served as lead for the conceptualization, original draft preparation and writing - review and editing. Marco Scutari and Richard J. McNally contributed equally to writing - review and editing.

The data and the code accompanying this tutorial are available in our repository on the Open Science Framework (<https://osf.io/fne9m/>), with information related to the package versions used to perform the different analyses.

We thank Sacha Epskamp and Lourens Waldorp for their very insightful comments that allowed our article to greatly improve during the review process.

Correspondence concerning this article should be addressed to Giovanni Briganti, Department of Psychology, Harvard University, 1232 William James Hall, 33 Kirkland Street, Cambridge, MA 02138-2044, United States. Email: giovanni.briganti@hotmail.com

reasoning (Pearl, 2009; Pearl & Mackenzie, 2018). That is, they can ascertain both the direction and the magnitude of causal effects (Maathuis et al., 2018). BNs are defined by a *directed acyclic graph* (DAG) and by the joint probability distribution of the variables under investigation. The role of the DAG is to express the *conditional independence* relationships between the variables (nodes) by using *graphical separation* (Koller & Friedman, 2009; Lauritzen, 1996; Pearl, 1988; Scutari & Denis, 2021). In other words, if two variables are separated in the DAG by some other variables, they are independent in probability conditional on (that is, after controlling for) those other variables. Hence, the role of the probability distribution is to express the magnitude of the causal effects linking the variables that are not graphically separated. Because BNs contain only directed edges, they are ideal for modeling the admissible causal relationships in observational data, thereby complementing insights provided by partial correlation networks. BNs should not be confused with pairwise Markov random fields estimated with Bayesian methods: These have been recently introduced in the network psychopathology literature in both methodological and applied research (Briganti et al., 2020; Williams & Mulder, 2020a, 2020b). Using BNs for rigorous causal inference requires several strong assumptions (Maathuis et al., 2018), such as the existence of a DAG underlying the data; *sufficiency*, that is, all the causes of a given variable are measured (which means that there are no latent variables and there is no selection bias); and *faithfulness*, that is, all the variables that are connected in a given way in the network are probabilistically dependent. To augment probabilistic inference with causal meaning the faithfulness and sufficiency assumptions need to be further strengthened into causal faithfulness and causal sufficiency, which will be discussed in more detail (see Structure Learning and Assumption Needed for Structure Learning sections). These assumptions of having an underlying DAG and causal sufficiency are difficult to verify in psychological data: For instance, we do not know whether symptoms from a checklist or psychometric measure or structured diagnostic interview assessing a mental disorder interact as a DAG except in very specific cases, such as the relationship between a traumatic event (the cause) and posttraumatic stress disorder (PTSD) symptoms (McNally et al., 2017). In most cases, such as depression, we do not know whether a given symptom unilaterally influences another (as in a DAG), whether two symptoms mutually influence each other (as in a partial correlation network), or whether there is one or more latent variables causing symptoms (as within the latent variable framework). The faithfulness assumption will likely be violated in a data set where the data generating process is such that $A \rightarrow B, B \rightarrow C, A \rightarrow C$, as the causal relationships are counterbalanced (Andersen, 2013; Cartwright & McMullin, 1984).

The insights that BNs can bring to the field of network psychopathology can be helpful in various situations, such as performing inference on the network structure based on its estimates. Edge weights are used, for instance, to identify the most central (i.e., interconnected) nodes in a network (Elliott et al., 2019): Summing edge weights (strength centrality) or estimating the shared variance R^2 (predictability) are the two most common ways of gauging the importance of symptoms (Boccaletti et al.,

2006; Haslbeck & Fried, 2017). Central symptoms are highly connected, with many connections or several strong ones. If such symptoms are the source of activation that maintains an episode of disorder, then successfully reducing their severity ought to cause a beneficial therapeutic cascade resulting in recovery from illness. Accordingly, they would be ideal therapeutic targets. Unfortunately, because the edges in partial correlation networks are undirected, it is impossible to tell whether a high-centrality symptom is the source of activation or the recipient of activation. Only if it is the source does it make sense to consider it a high-priority target for clinical intervention. Yet because they have often been theorized as potential targets for clinical intervention, centrality estimates are defined to operate on undirected networks and thus do not take edge directions into consideration. In other words, they do not ascertain whether central symptoms *cause* or *are caused by* many symptoms in the network; if a symptom is well-connected because it is highly *controllable* (that is, caused by several symptoms connected to it), acting on it is unlikely to affect other symptoms. Because interpreting network structures is an important part of any network study, complementing the study of interaction among symptoms with further causal insight is essential.

Only a handful of studies have applied BNs in psychopathology research. These studies have included dissociation (Cernis et al., 2021), self-harm (Hinze et al., 2021), paranoia (Bird et al., 2019), psychosis (Kuipers et al., 2019; Moffa et al., 2017), obsessive-compulsive disorder (Jones et al., 2018; McNally et al., 2017), bipolar disorder (McNally et al., 2021), rumination (Bernstein et al., 2017), depression (Briganti et al., 2021), PTSD (McNally et al., 2017), and alexithymia (Briganti et al., 2020). In these studies, BNs illuminated potential causal relations among symptoms in cross-sectional data sets. However, the authors often varied in their computational methods, thereby underscoring the need for an accessible tutorial reference for researchers keen to apply BNs to study mental disorders. The present work aims to fill this gap by introducing BNs, by providing key commands in the R environment for statistical computing (R Core Team, 2020) to estimate BNs from data, and by addressing the main strengths and limitations of BNs in psychopathology.

This tutorial builds on previous empirical studies that applied the two main approaches for estimating BNs: constraint-based algorithms (Briganti et al., 2020, 2021) and score-based (McNally et al., 2017, 2017; Moffa et al., 2017). This tutorial complements these applied works by offering a principled introduction to BNs and on how to interpret them in psychological data.

This article is structured as follows. First, we will introduce BNs and their main theoretical properties. We will then outline both constraint-based and score-based structure learning algorithms as well as their implementation in R, illustrating them with data and code. Finally, we will discuss the interpretation of BNs in observational studies as well as answer common questions researchers may have about BNs. More extended exposition on these topics appear in specialized textbooks (Koller & Friedman, 2009; Neapolitan, 2004; Pearl, 2009; Scutari & Denis, 2021) and in an accessible trade book for the general reader (Pearl & Mackenzie, 2018).

Bayesian Networks

Definition

BNs are defined as the combination of a network structure, specifically a DAG, and a probability distribution. The DAG provides a high-level abstraction useful for qualitative reasoning in the context of exploratory analysis and for testing hypotheses about how symptoms relate to one another. The probability distributions of the symptoms quantify those relationships in terms of their magnitude and direction.

Directed Acyclic Graphs

The network structure of a BN is a mathematical object $\mathcal{G} = (V, A)$ consisting of a set of *nodes* $V = \{v_1, v_2, \dots, v_N\}$ (where N is the number of nodes) and of a set of edges or arcs $A = \{(v_1, v_2), (v_2, v_3), \dots\}$ that represent all connections between two nodes. Given V , a graph is uniquely identified by A . In the case of a DAG, the edges can only be directed, that is, $(v_i, v_j) \neq (v_j, v_i), v_i \rightarrow v_j$, with the assumption that two nodes are only connected by one edge. In each edge $v_i \rightarrow v_j$, v_i is called the “parent” node and v_j is called the “child” node. DAGs are also acyclic: They do not contain any loops, that is, edges from a node to itself $v_i \rightarrow v_i$; or any cycle, that is, sequences of edges that start and end on the same node, such as $v_i \rightarrow v_j \rightarrow \dots \rightarrow v_k \rightarrow v_i$. The primary role of the DAG in a BN is to express the set of conditional independence of relationships among variables (that is, variables that do not cause each other).

Markov Blankets

The DAG of a BN is an independence map of the probability distribution of the variables X_i , which are represented in the networks by the nodes v_i ; retrieving such a map means determining which nodes are conditionally (in)dependent. This is achieved by using d -separation, which defines graphical separation for DAGs and provides a way to algorithmically determine whether two nodes in a network are (in)dependent or conditionally (in)dependent (Geiger et al., 1990). Formally, two nodes v_i and v_j are d -separated by a conditioning set of nodes S if conditioning on all members of S block all paths (sequence of nodes and edges with v_i as starting node and v_j as ending node, that do not necessarily follow the direction indicated by edges) between v_i and v_j . A collider does not generate an unconditional association between the variables that determine it (Greenland et al., 1999; Pearl, 1988). In Figure 1, three examples of a basic three-node DAG are illustrated, as well as how they differ in conditional independences.

The set S that makes a node v_i independent of all other nodes in G is known as the Markov blanket of v_i . By the Markov property, the Markov blanket includes a node’s parents, children, and spouses, that is, children’s other parents. The Markov blanket is useful in investigating a target node of interest while ignoring the rest of the BN; all nodes outside of the Markov blanket are independent from the node of interest after controlling for those in the Markov blanket itself.

The Markov Property

If two nodes are not connected by an edge in a BN, then they are either independent or conditionally independent given some other nodes: this is called the local Markov property (Korb & Nicholson, 2010; Lauritzen, 1996). The global Markov property is a stronger version of the local property and generalizes it to sets of variables: any two subsets of variables are conditionally independent given a separating subset (Föllmer, 1980; Lauritzen, 1996). Graphical separation implies probabilistic independence¹

$$v_i \perp\!\!\!\perp_G v_j | v_k \Rightarrow v_i \perp\!\!\!\perp_P v_j | v_k$$

(where $\perp\!\!\!\perp_G$ means graphical separation and $\perp\!\!\!\perp_P$ probabilistic independence), making the network itself a clear representation of the conditional independence relationships between nodes. For this reason, the DAG is called an independence map of the variables.

The Markov property makes it possible to write²

$$Pr(\mathbf{X}, \Theta) = \prod_{i=1}^N Pr(X_i | \Pi_{X_i}; \Theta_{X_i}),$$

decomposing the larger model $Pr(\mathbf{X}, \Theta)$ into a set of smaller models $Pr(X_i | \Pi_{X_i}; \Theta_{X_i})$, one for each variable X_i conditional on its parents Π_{X_i} in the DAG (Korb & Nicholson, 2010), Θ and Θ_{X_i} denote the parameters of the respective models. These smaller models are individually much simpler than $Pr(\mathbf{X}, \Theta)$; and they are topical, in the sense that they focus on a single variable and its relations with variables that surround it in the DAG. This factorization follows from the Markov property of BNs, and it is only possible because of the absence of loops and cycles in the graph. Figure 2 represents a toy example Bayesian Network of four nodes from specific manic symptoms from the Young Mania Rating Scale (Young et al., 1978), and an example of how a probability distribution in a DAG is factorized (Briganti et al., 2021).

Figure 2 also shows a fundamental type of relationship in a BN, represented by the three nodes increased energy, language-thought disorder, and speech disorder. Both increased energy and language-thought disorder have an edge pointing to speech disorder and they are not adjacent, that is, they do not share an edge with one another. This pattern of edges is commonly known as a v -structure, or a collider, and it is one of the building blocks of BNs. In a collider ($A \rightarrow C \leftarrow B$), four conditional independences are found: A and C are dependent, B and C are dependent, A and B are independent, and A and B are dependent conditional on C. This means that two independent variables, A and B, can become dependent when one conditions on C, because observing the common consequences of two independent causes can modify their probability (Pearl, 2009). The two causes (A and B) are therefore negatively partially correlated (if they are both positive or both negative), which is counterintuitive; conditioning on the common

¹ If two nodes v_i and v_j are separated in the graph, then their probabilities are also independent.

² The joint probability of the set of variables X and its parameters is equal to the product of probabilities of each of the smaller models $Pr(X_i | \Pi_{X_i}; \Theta_{X_i})$ for each variable X_i given its parents Π_{X_i} and its parameters Θ_{X_i} .

Figure 1
Three Examples of DAGs

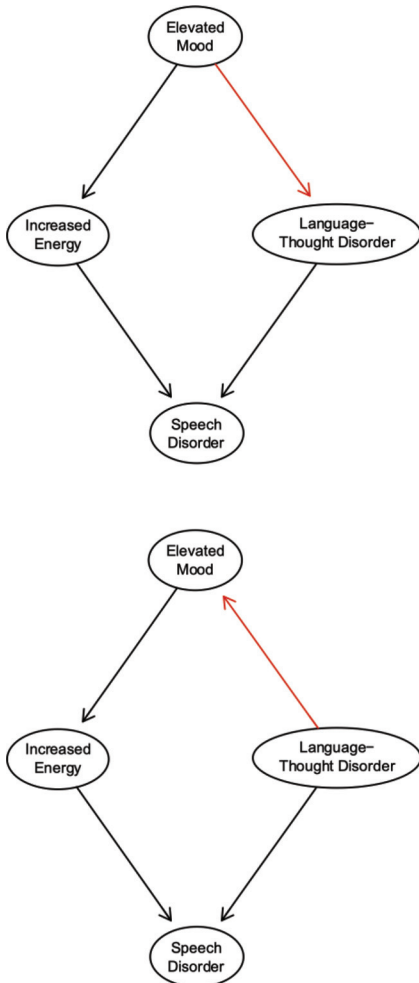


Note. The DAG on the left is also called a chain. The DAG in the middle is called a common cause model (Z is the common cause of X and Y). The DAG on the right is called a collider (X and Y cause Z). In the DAG on the left and the one in the middle, X and Y are not independent, but they are conditionally independent given Z. In the DAG on the right, X and Y are independent, but they are not conditionally independent given Z.

effect (C) in the collider (that is, studying the associations while conditioning on the effect) leads to different estimates compared to studying the two causes individually. This phenomenon is known as collider bias (Berkson, 1946), and in network analysis it

is known to induce spurious edges among nodes, and likely affects the generalizability of results when applied to the wrong population (De Ron et al., 2019; McNally, 2021). DAGs are tools that can help researchers identify instances of collider bias.

Figure 2
An Example of Bayesian Network of Four Nodes



$$\Pr(\text{Elevated Mood, Increased Energy, Language-Thought Disorder, Speech Disorder}) = \left. \begin{array}{l} \text{Joint} \\ \text{distribution} \end{array} \right\}$$

$$\begin{aligned} &\Pr(\text{Elevated Mood}) \\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood}) \\ &\Pr(\text{Language-Thought Disorder} \mid \text{Increased Energy, Elevated Mood}) \\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder, Increased Energy, Elevated Mood}) = \end{aligned} \left. \begin{array}{l} \text{Chain rule} \end{array} \right\}$$

$$\begin{aligned} &\Pr(\text{Elevated Mood}) \\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood}) \\ &\Pr(\text{Language-Thought Disorder} \mid \text{Elevated Mood}) \\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder, Increased Energy}) = \end{aligned} \left. \begin{array}{l} \text{Bayesian} \\ \text{network,} \\ \text{local} \\ \text{distributions} \end{array} \right\}$$

$$\begin{aligned} &\Pr(\text{Elevated Mood}) \\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood}) \\ &\Pr(\text{Language-Thought Disorder, Elevated Mood}) \\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder, Increased Energy}) = \end{aligned} \left. \begin{array}{l} \text{Bayes} \\ \text{theorem} \end{array} \right\}$$

$$\begin{aligned} &\Pr(\text{Increased Energy} \mid \text{Elevated Mood}) \\ &\Pr(\text{Language-Thought Disorder, Elevated Mood}) \\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder, Increased Energy}) = \end{aligned} \left. \begin{array}{l} \text{Reversing} \\ \text{a score-} \\ \text{equivalent arc} \end{array} \right\}$$

$$\begin{aligned} &\Pr(\text{Increased Energy} \mid \text{Elevated Mood}) \\ &\Pr(\text{Elevated Mood} \mid \text{Language-Thought Disorder}) \\ &\Pr(\text{Language-Thought Disorder}) \\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder, Increased Energy}) \end{aligned}$$

Note. Elevated Mood has two directed edges towards increased energy and language-thought disorder, and speech-disorder receives two directed edges from increased energy and language-thought disorder. Increased energy and language-thought disorder are d-separated given elevated mood and speech disorder. Elevated mood and speech disorder constitute the Markov blanket of both increased energy and language-thought disorder. Increased energy, language-thought disorder and speech disorder constitute a v-structure (collider). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Equivalence Classes

Although a factorization of the probability distribution of X is uniquely identified by a DAG, a DAG is not uniquely identified by a factorization of the probability distribution. For instance, patterns of edges like $v_j \rightarrow v_i \rightarrow v_k$, $v_j \leftarrow v_i \leftarrow v_k$ and $v_j \leftarrow v_i \rightarrow v_k$ lead to equivalent probability distributions (see Appendix).

Hence, several edges in a DAG can be reversed without changing the conditional (in)dependence relationships in X . In other words, if no new v-structure (or cycle) is created, we are just producing equivalent DAG representations of the same set of conditional (in)dependence relationships. All such DAGs are then part of an equivalence class that is uniquely identified by the undirected graph underlying these DAGs and by the v-structures. The direction of other edges is either uniquely identified because one of two directions would introduce a new v-structure or a cycle (these are sometimes called compelled edges), or completely undetermined. This results into a completed partially directed graph (Castelletti et al., 2018; Chickering, 2002a).

Probability Distributions for Bayesian Networks

In this section, we introduce the most common probability distributions used in BNs as well as their limitations.

The three most common probability distributions for BNs are discrete, ordinal, Gaussian, and conditional linear Gaussian (Scutari & Denis, 2021). These probability distributions satisfy the following requirements: First, the structure of the BN should be learned efficiently from data by using statistical tests or goodness-of-fit measures; second, the BN should be flexible, that is, distributional assumptions should not be too strict; third, the BN should be easy to query to perform inference.

Discrete and Ordinal Bayesian Networks

In discrete BNs, local distributions of $X_i \mid \Pi_{X_i}$ are composed of conditional probability tables for each node given all other parent nodes, that is, the configuration of values of its parents (Spirtes & Meek, 1995). Learning the structure of BNs from discrete variables is particularly useful in the case of psychological data because most measurement tools are scored on Likert scales. Although using discrete variables is a helpful solution so that dependencies need not be linear, there may be some loss of information, particularly in the case of ordinal variables (such with Likert scales) that will be treated as categorical, because the ordinality of data is lost, as shown in the related domain of GGMs (Isvoranu & Epskamp, 2021). Moreover, to deal with this problem, many applied researchers treat Likert data as continuous in estimating network models (Briganti et al., 2018, 2019). This is why investigators have tested how well one can estimate BNs from ordinal data (Cui et al., 2016; Luo et al., 2020; Musella, 2013; Tsagris et al., 2018).

Gaussian Bayesian Networks

Gaussian BNs (which are different from Bayesian GGMs) follow a multivariate normal distribution with mean μ and a variance-covariance matrix Σ , $X \sim N(\mu, \Sigma)$, with a precision matrix Ω computed as the inverse of the variance-covariance matrix $\Omega = \Sigma^{-1}$.

The partial correlation coefficients ρ_{ij} , which model the residual correlation between two nodes after controlling all other nodes, are computed from the precision matrix as $\rho_{ij} = -\Omega_{ij} / \sqrt{\Omega_{ii}\Omega_{jj}}$ where Ω_{ij} is the element of Ω . If v_i and v_j are d-separated in the DAG (or in the case of violation of the faithfulness assumption), the absence of an edge $v_i \rightarrow v_j$ in the DAG between the two nodes by definition implies the conditional independence ($\rho_{ij} = 0$) between the two variables X_i and X_j given all subsets of the remaining variables (we refer the reader to The Inductive Causation Algorithm section to have a better understanding on why going through all subsets of variables is necessary). From this, it follows that most edges of a Gaussian BN belong to a subspace of a GGM (a partial correlation network), except specific independence relationships, such as a collider structure. On the other hand, cycles can be represented in a GGM but not a Gaussian BN.

As for the local distributions, $X_i \mid \Pi_{X_i}$ are linear regression models with Π_{X_i} acting as regressors. This assumes that all dependencies are linear (Grzegorzczuk, 2010), a requisite for the distribution of X to be multivariate normal. This may be advantageous in modeling Likert scales, but only if these scales are designed so that the items on the scale increase in linear increments (that is, Item 2 implies twice the intensity of Item 1, Item 3 implies three times the intensity of Item 1, and so forth). When this is not the case, the numbering of the items on the Likert scale leads the BN to incorrectly interpret the items as equidistant from each other and the resulting estimates of the ρ_{ij} will be biased.

Gaussian BNs have two main limitations. First, few real-world multivariate data sets follow a multivariate Gaussian distribution (the same applies to partial correlation networks since the probability distribution is the same). Second, computing partial correlations can be challenging in data sets with a large number of variables because of singularities, that is, perfect linear relationships between two variables (Lin et al., 2014). Although exact linear correspondence rarely occurs, relationships that are almost singular may result in numerical precision issues and may make BNs difficult and time-consuming to learn.

Conditional Linear Gaussian Bayesian Networks

Conditional linear gaussian BNs contain both discrete and continuous nodes and subsume both Discrete BNs and Gaussian BNs. In terms of local distributions, discrete nodes have conditional probability tables, whereas continuous nodes have a set of linear regression models, one for each configuration of the discrete parents, and with the continuous children as regressors (Cowell, 2005). In such networks, continuous nodes cannot be parents of discrete nodes, which is a strong limitation for treating mixed data: only a given set of edges can exist in the DAG, which is an important limitation on what models can be encoded by the network structure.

Structure Learning of Bayesian Networks

Structure learning algorithms learn the structure of BNs from data, possibly integrating external expert knowledge if available. Constraint-based algorithms use statistical tests to learn conditional independence relationships (the constraints themselves); score-based algorithms rank candidate DAGs based on some goodness-of-fit criterion; hybrid algorithms use conditional independence

tests to exclude most candidate DAGs, and then perform a score-based search on those that are still under consideration. In addition, specific edges can be whitelisted (included) or blacklisted (excluded) in candidate DAGs by default based on prior knowledge: this helps the structure learning algorithm in its choice of good candidate models (Scutari & Denis, 2021).

All algorithms in this section are implemented in the R package `bnlearn` (Scutari, 2010), Version 4.7. Two other packages that implement some algorithms for BNs are `pcalg` (Kalisch et al., 2012; Kalisch & Bühlmann, 2007) and `BiDAG` (Suter et al., 2021). To showcase the functions and their output, we use a data set consisting of seven key items from the 20-item Zung Self-Rating Depression Scale (Zung, 1965) from 1,090 subjects. A detailed description of the data appears elsewhere (Briganti et al., 2021). In this case, we selected only seven nodes out of the 20 nodes in the scale to simplify the data visualization in the tutorial. The data and the code accompanying this tutorial are available in our repository on the Open Science Framework (<https://osf.io/fne9m/>; Briganti, 2021), with information related to the package versions used to perform the different analyses (Briganti et al., 2021). To read the data, one can use the following commands

```
library("readr")

data <- read_table("data.csv", sep=";", header=TRUE)

# this loads the data
```

Assumptions in Structure Learning

Most structure learning algorithms make several fundamental assumptions (Dawid, 2010; De Campos & Ji, 2011; Maathuis et al., 2018):

- The underlying structure that is to be learned is indeed a DAG.
- There is no selection bias and there are no latent or confounding variables, that is, all the common causes of all measured variables have been measured: this assumption is also known as causal sufficiency.
- Only the variables that are *d*-separated in a DAG will be independent, the others will be dependent: This assumption is also known as causal faithfulness, which is the converse of the Markov property.
- The only relationships between the random variables in X should be conditional independencies, the only kind of relationships that can be encoded by BNs.
- Each node in the network must represent one random variable in X : There should not be multiple nodes which are deterministic functions of the same random variable in X (such as a sum score of two variables). This is usually the case in psychological networks since separate symptoms are used as input.
- Observations must be independent realizations; if not, the network should be otherwise defined, such as a dynamic BN in the case of temporal dependence, because observations are not independent as a single subject is measured over multiple occasions (Song et al., 2009).

- The global probability distribution $P(X)$ must be strictly positive: that is, every combination of every possible value of the variables in X must represent an event that is observable in principle; this is needed to have uniquely determined Markov blankets and therefore an identifiable model. This is always the case in psychological data, because we measure events that patients report (e.g., fatigue) or that we observe (e.g., crying).

Most, but not all, algorithms require satisfaction of these assumptions, and some are incapable of testing whether the assumptions hold. Algorithms in the fast causal inference (FCI) family are consistent when selection bias occurs or when a latent common cause is present. Using this kind of approach has multiple implications for drawing inferences. For instance, a directed edge from Node A to Node B in the presence of latent variables entails that A is a cause of B, but does not entail that A is a direct cause of B (Maathuis et al., 2018).

For examples of data sets where the assumption of independent realizations is violated due to temporal dependencies among variables, see studies on time-series networks (Bringmann et al., 2013; Epskamp, 2020) or panel data networks (Briganti et al., 2021).

Stability in Structure Learning Algorithms

The structure of BNs may vary when performing a learning algorithm several times. For cross-sectional studies, it is desirable to study a “stable” network structure, that is, a set of edges and directions that is unlikely to vary. Stability is an important topic in network analysis: researchers need to study stable networks whose structure is likely to be influenced by factors such as sample size. For partial correlation networks, one can use bootstrapping methods to evaluate the stability of network estimates (Epskamp et al., 2018). It is possible to directly account for stability in structure learning in a similar way: The same algorithm learns a sufficiently large number of BNs from bootstrap samples and we only consider edges that appear in a proportion of BNs higher than some threshold. Previous empirical papers (Briganti et al., 2020, 2021) only included edges that appeared in more than 85% of networks (this is called strength), and whose direction appeared in more than 50% of networks (this is called minimum direction). We recommend that researchers only report stable BNs obtained in this way, and that they should use 100–200 bootstrap samples to ensure the proportion of BNs in which each edge appears are estimated accurately. In addition, the proportion of times each edge was included in the stable BNs can be reported.

Constraint-Based Algorithms

Constraint-based algorithms are based on Pearl and Verma’s (1991) work on causal graphical models. This class of algorithms identifies conditional independencies (the “constraints”) with statistical tests and connects nodes that are not independent. The algorithms determine the sequence of tests that will be applied to the data, typically starting from marginal tests (that is, not conditioning on any variable) and gradually increasing their complexity (adding conditioning variables) to maximize speed and statistical power while ensuring the accuracy of the learned networks.

Tests for Conditional Independence

Two random variables X and Y are conditionally independent given a third random variable Z if, given knowledge of Z , knowledge of whether X occurs provides no information on the likelihood of Y occurring, and knowledge of whether Y occurs provides no information on the likelihood of X occurring (Dawid, 1979). Testing for conditional independence is a necessary step for constraint-based algorithms because conditional independence relationships themselves constitute the constraint. We hereby cite three default conditional independence test in `bnlearn`; further details appear in comprehensive reviews of the topic (Edwards, 2012). The asymptotic χ^2 mutual information test is the default for categorical and mixed variables: It is an information-theoretic distance measure related to the deviance of the tested models and proportional to the log-likelihood ratio (the two differ by a factor), with the degrees of freedom equal to the number of free parameters in the two models (under the assumption that all parameters can be estimated). The Hotelling's exact test is the default for continuous variables: it is the multivariate version of the Student's t for testing the null hypothesis that the correlation is null. The Jonckheere-Terpstra test is the default for ordinal variables: it is a trend test that generalizes the Wilcoxon test (which only tolerates one ordinal variable compared to a binary factor) to two ordinal variables. A comprehensive review of methods to test for conditional independence exceeds the scope of this tutorial, but appears in the `bnlearn` manual (Scutari, 2010).

The Inductive Causation Algorithm

The inductive causation (IC) algorithm (Pearl & Verma, 1991) was the first and simplest structure learning constraint-based algorithm. Although more modern algorithms are preferable for practical applications, all evolved from IC and thus share the same fundamental steps; hence we will describe IC in detail for illustrative purposes.

The algorithm starts from a complete graph in which each variable is connected to all other variables by an undirected edge. First, for each pair of variables A and B in \mathbf{X} , the algorithm searches for a set S_{AB} such that A and B are independent given S_{AB} and such that A and B are not part of it ($A, B \notin S_{AB}$). Edges between all such pair of variables are removed. The algorithm composes S_{AB} starting from a list of no variables (with an empty S_{AB}) then each single variable other than A and B in \mathbf{X} , then with a combination of other variables, until all combinations have been tried to remove the edge. Because of this, algorithms derived from the IC have one weakness: the complexity in the number of nodes in a graph (Li et al., 2019). Second, for each pair of variables that are not connected by an edge but are both connected to a common neighbor C , the algorithm checks whether $C \in S_{AB}$: if $C \notin S_{AB}$, then the direction of edge $A - C$ becomes $A \rightarrow C$ and that of edge $C - B$ becomes $C \leftarrow B$. Third, the direction of the edges that are still undirected is set following two rules: if A is adjacent to B and there is a strictly directed path from A to B (that is, all edges in the path are directed), then $A - B$ becomes $A \rightarrow B$; if A and B are not adjacent but $A \rightarrow C$ and $C - B$, then $C - B$ becomes $C \rightarrow B$. This step ends the algorithm which then returns the completed partially directed graph in which only those edge directions that can be uniquely identified from the data are represented.

The IC algorithm is implemented by the Peter and Clark (PC) algorithm (Spirtes et al., 1993), which starts from a saturated network and then performs tests that gradually increase the number of conditioning nodes. Constraint based algorithms treat BNs as perfect maps,³ that is

$$A \perp\!\!\!\perp_P B \mid C \Rightarrow A \perp\!\!\!\perp_G B \mid C.$$

This assumption is also called faithfulness, and it is analogous to causal faithfulness from a purely probabilistic perspective. This assumption cannot be empirically verified. However, the assumption has the advantage of not requiring a strictly positive $P(X)$ to have uniquely defined Markov blankets.

In the R package `bnlearn` Version 4.7 (Scutari, 2010), the PC algorithm is implemented in the `pc.stable()` function. To plot the graph in a consistent fashion with all other network analyses in psychological literature, we use the `qgraph` package Version 1.9 (Epskamp et al., 2012).

First, we load the necessary packages, that is, `bnlearn` and `qgraph`.

```
library("bnlearn") # this loads the bnlearn package
library("qgraph") # this loads the qgraph package
```

Second, we perform the PC algorithm to learn the structure of the BN.

```
BNpc <- pc.stable(data) # this performs the algorithm
qgraph(BNpc) # this plots the network
```

Third, we obtain a stable BN.

```
BST <- boot.strength(data, # this includes the data set
  R = 200, # this sets the number of boots
  algorithm = "pc.stable") # this sets the algorithm
# note that boot.strength is likely to give warnings
```

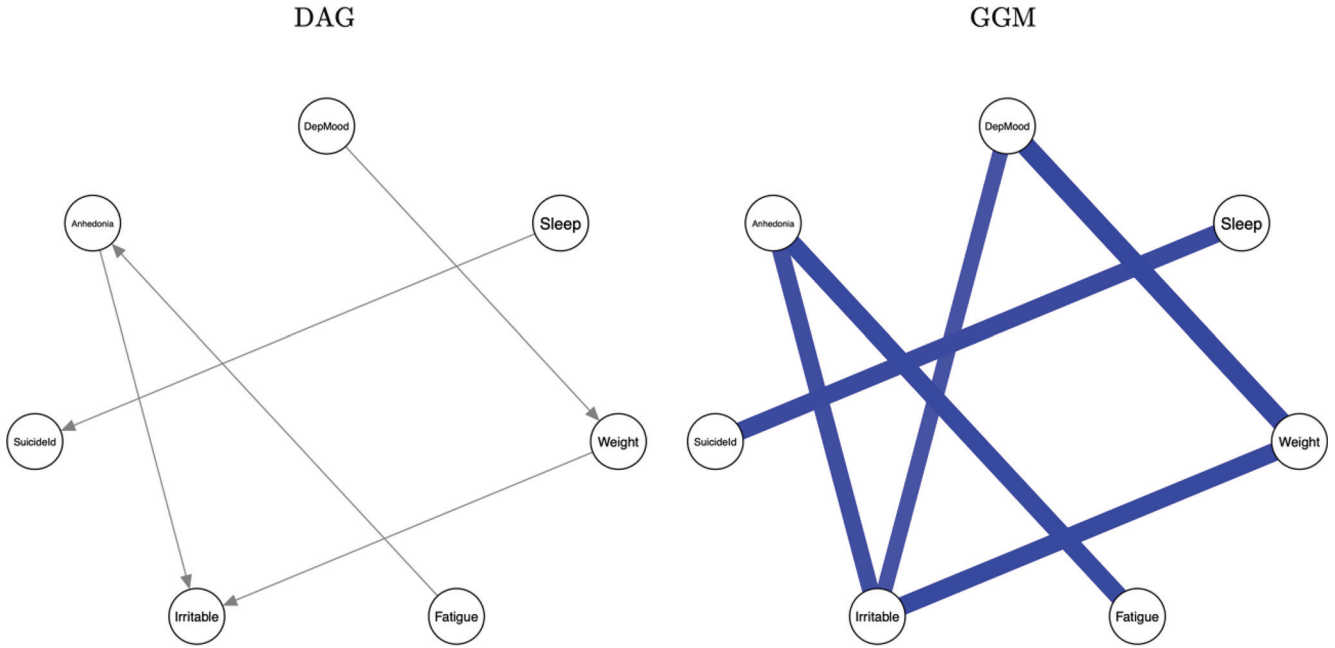
Fourth, we include the edges that appear in 85% of networks and whose direction appears in more than 50% of networks, and we compute the averaged network.

```
bst1 <- BST[BST$strength > 0.85 & # this sets the strength
  BST$direction > 0.5,] # this sets the minimum
  direction
avnet1 <- averaged.network(BST,
  threshold = 0.85) ## compute the average
  network
```

³The probabilistic independence of A and B given C is true if and only if the graphical independence of A and B is true.

Figure 3

A BN Learned Through the PC Algorithm (Left) Compared With a Regularized Partial Correlation Network (Right)



Note. Edge thickness and color saturation denote the edge weight. See the online article for the color version of this figure.

Finally, we plot the averaged network with qgraph.

```
qgraph(avgnnet1, layout="circle") # this plots
  the network with qgraph
```

The resulting network can be visualized in Figure 3.

Other Constraint-Based Algorithms

Other well-known constraint-based algorithms include the grow-shrink (GS) algorithm (Margaritis, 2003) and the incremental association (IAMB) sets of algorithms (Tsamardinos et al., 2003), which first learn the Markov blanket for each node in the network to reduce the number of tests required by the IC algorithm. The initial network is assumed not to have any edges.

```
BNgs <- gs(data) # this performs the GS algorithm
```

```
BNiamb <- iamb(data) # this performs the IAMB algorithm
```

Score-Based Algorithms

Score-based structure learning algorithms assign a score to each candidate BN. Network scores focus on the DAG as a whole (as opposed to individual constraints) and are goodness-of-fit statistics that assess how well the DAG mirrors the dependence structure in the data (Scutari & Denis, 2021). Two methods for scoring are the Bayesian information criterion (BIC) and posterior probabilities from flat priors (Daly et al., 2011; Friedman & Koller, 2003; Geiger & Heckerman, 2002; Goudie & Mukherjee, 2016; Grzegorzcyk & Husmeier, 2008; Heckerman & Geiger, 1995; Kuipers et

al., 2018b; Madigan et al., 1995; Pensar et al., 2020). Score-based algorithms are typically applications of heuristic search algorithms, which are not guaranteed to return the highest-scoring DAG. Exact search algorithms, which do provide this guarantee, are markedly slower than heuristic algorithms and are not feasible for many practical applications (Cussens, 2012; Koivisto & Sood, 2004; Scanagatta et al., 2015; Suzuki, 2017). Hence, we will not consider them further.

The Hill-Climbing Algorithm

The Hill-Climbing (HC) algorithm (Russell & Norvig, 2002) is a greedy search algorithm that explores DAGs by single-edge additions, removals, and reversals. First, the algorithm chooses an initial candidate network structure \mathcal{G} (it chooses by default the empty DAG) and computes its score $Score(\mathcal{G})$. Second, it sets $maxscore = Score(\mathcal{G})$.⁴ Third, the algorithm computes the score of the modified networks \mathcal{G}^* obtained by every possible edge addition, deletion or reversal that does not result in a cyclic network. If any \mathcal{G}^* has $Score(\mathcal{G}^*) > Score(\mathcal{G})$, the new score is updated in $maxscore = Score(\mathcal{G}^*)$ and \mathcal{G}^* becomes the new candidate network. Finally, the algorithm returns the final network when no \mathcal{G}^* has a higher score than \mathcal{G} .

```
BNhc <- hc(data) # this performs the HC algorithm
```

⁴ $maxscore$ represents the maximum score computed from a network structure; the goal of the algorithm is to find a modified network structure that will have a higher score than $maxscore$, then updates $maxscore$ as it goes, until it finds the highest score possible.

Other Score-Based Algorithms

Three well-known score-based algorithms include the greedy equivalence search (Chickering, 2002b), a HC algorithm over equivalence classes instead of graphs (which minimizes the search space), the tabu search (Glover, 1990), a modified HC algorithm that does not stop at the first DAG for which every possible edge addition, deletion, or reversal does not improve the score, but further explores the space of DAGs to find a better DAG; and genetic algorithms (inspired by evolutionary biology) that perturb (mutation) and combine (crossover) features through several generations of structures while retaining those yielding better scores (Davis, 1991).

```
BNtabu <- tabu(data) # this performs the Tabu algorithm.
```

Hybrid Algorithms

Hybrid algorithms combine the two previous classes of constraint-based and score-based algorithms: They are the state of the art for many problems (Scutari & Denis, 2021). They work by learning some conditional independence constraints to reduce the number of candidate networks prior to finding the network that maximizes a given score function (Kuipers et al., 2018a; Tsamardinos et al., 2006).

The sparse candidate algorithm (Friedman et al., 2013) was the first example of this approach in the literature. It is an iterative algorithm that restricts the parents of each variable to belong to a small subset of candidates. First, in a network \mathcal{G} , the algorithm restricts the candidate parents for a node $X_i \in \mathbf{X}$ to the parents of X_i in \mathcal{G} in a set \mathcal{C}_i . Second, the algorithm finds the network structure \mathcal{G}^* that maximizes $Score(\mathcal{G}^*)$ among the networks in which the parents of each node X_i are included in the corresponding set \mathcal{C}_i , then sets the new network $\mathcal{G} = \mathcal{G}^*$. After the steps are repeated, the algorithm returns the graph \mathcal{G} .

The sparse candidate algorithm is implemented in the `rsmax2()` function, which performs the step of restricting and maximizing only once as further steps rarely yield better DAGs (Friedman et al., 2013).

```
BNrs <- rsmax2(data) # this computes the rsmax2 algorithm.
```

Common Issues to Consider When Learning BNs

Choice of Algorithm

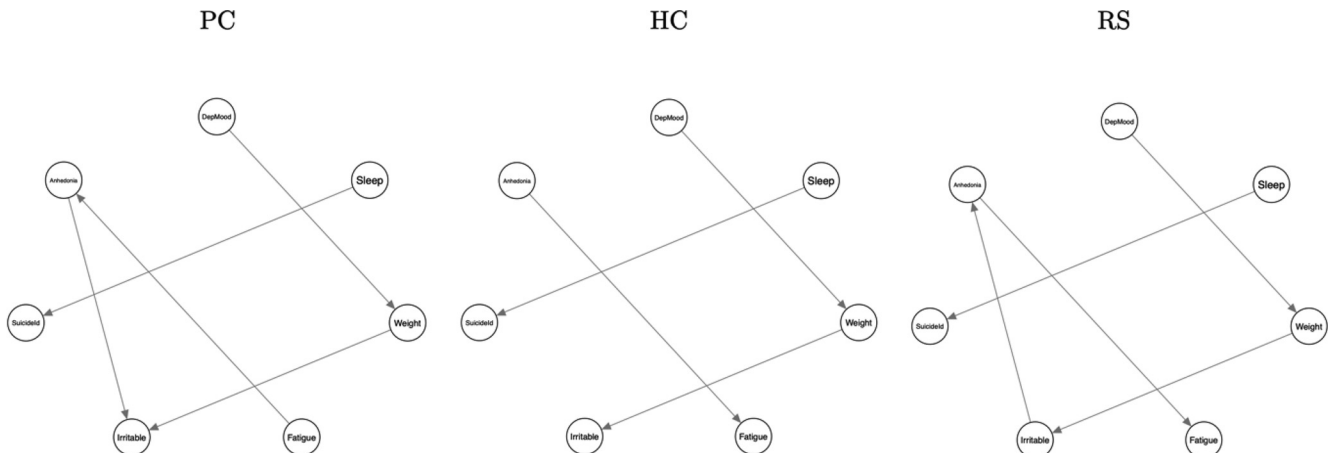
Which approach should psychopathologists choose to learn the structure of a BN? The number of possible configurations of algorithms, scores, conditional independence tests, and all the tuning parameters that accompany them is indeed overwhelming. A comprehensive simulation study showed that there were no systematic differences in performance or sensitivity to error in real-world data when comparing the three classes of learning algorithms (Scutari et al., 2018). If we consider individual algorithms, none consistently outperforms the others discussed here. This is also the case in our own example data set: The three classes of algorithms retrieve very similar structures, albeit with some differences in edge directions (see Figure 4). Another comprehensive comparison of different structure learning algorithms showed that choosing an algorithm also depends on its reliability and resilience, and that synthetic performance may overestimate real-world performance (Constantinou et al., 2021).

Although researchers should clearly present the algorithm chosen for structure learning and report the stable network structures, we recommend that they also perform the two other classes of structure learning algorithms as robustness checks. This would disclose the edge directions that vary among algorithms.

Assumptions Needed for Causal Inference

BNs whose DAGs are in the same equivalence class are probabilistically indistinguishable: It is impossible to choose one over the other from data without resorting to additional expert knowledge. Therefore, inferring causal relationships from observational data should be done very carefully: The causal effects whose direction is supported by the data are those that correspond to edges with a defined direction in the completed partial DAG for the equivalence class.

Figure 4
BNs Learned With the Constraint-Based PC Algorithm (Left), the Score-Based HC Algorithm (Middle), and the Hybrid RS Algorithm (Right)



Furthermore, to defend a causal interpretation of a BN we need to satisfy three assumptions. First, each variable X_i is conditionally independent of its direct and indirect effects and from other variables given its direct causes: this is known as the causal Markov assumption.

Second, there is a DAG faithful to the probability distribution of X such that the dependencies arising from d-separations in the DAG are the only ones in the probability distribution. The faithfulness assumption for causal inference implies that any population produced by the DAG has the same conditional independence relationships obtained by applying d-separation to it. However, this assumption can be challenged by selection bias; that is, the population can be “unfaithful” to the DAG. When estimating DAGs representing the causal relationships among symptoms of mental disorders or psychological constructs, the population could be unfaithful to the DAG if it were estimated from a specific subsample. For example, a network structure of depressive symptoms in manic patients will show different conditional independence relationships than a network structure estimated from unipolar depressive patients. For further details concerning selection bias in network psychopathology, we refer to more complete overviews of the topic (De Ron et al., 2019; McNally, 2021).

Third, there must be no latent variable that acts as a confounding factor thus introducing spurious dependencies among the observed variables, which translate to spurious arcs in the BN. A confounding variable A is such that, in a path $A \rightarrow B \rightarrow C$, $A \rightarrow C$: this is also known as a backdoor path from A to C . In psychological data sets, confounding variables are an important issue since in many cases all relevant variables are not measured or included in a model. This assumption also concerns many studies designed from a latent variable point of view whereby items or symptoms presumptively reflect unobserved factors that cause their emergence and covariance. A possible solution to help overcome this issue is eliminating topological overlap among nodes (Fried & Cramer, 2017). That is, we can create composite variables to ensure that items tapping the same construct do not appear as individual, near-synonymous nodes. Hence, each node would constitute a composite variable representing a domain. This has been done with both gaussian graphical models and BNs (Briganti et al., 2019, 2020; Briganti & Linkowski, 2019). Another way of reducing the number of network components is by using the goldbricker method in the networktools package (Jones, 2017) that identifies colinear network components (that is, topologically overlapping items).

These assumptions are difficult to verify in real-world settings, and the best solution to tackle them is by using blocking in a carefully planned experimental design to screen out confounding factors via randomization, which severs any incoming causal link between the randomized variables and possible exogenous effects, but it is never applied to cross-sectional network studies. In the best-case scenario, scientific experiments can identify a small set of BNs that plausibly fit the data from a causal point of view.

What Kind of Data Are Needed for Learning?

The data used to learn the structure of a BN and to estimate its parameters are a key driver of the quality of the models we can obtain.

For structure learning, it is important that we have enough data to rank networks accurately using scores and to ensure that the conditional independence tests we use to assess constraints have sufficiently low rates of Type-I and Type-II errors. The ability of structure learning algorithms to recover from errors (that is, incorrectly including or excluding edges in the DAG) is limited, which makes it important to

limit the possibility of such errors occurring. For parameter learning, it is important to have enough data to detect the effect of the parents of each child and to estimate them with sufficient precision. In both cases, we may draw from existing literature on power and sample size calculations for linear and logistic regressions (Demidenko, 2007; Dupont & Plummer, 1998). When the sign of the effect associated with each edge is important, we may complement traditional power calculations by assessing Type-S errors (Gelman & Carlin, 2014). In both cases, it is important to remember that both structure and parameter learning operate at the level of the local distributions $X_i | \Pi_{X_i}$: we only need a sample size large enough to handle the most complex of them. If we assume that the network we are estimating is sparse, so that all nodes have a limited number of parents (say, less than 5), then the sample size required to learn a BN does not necessarily scale with the number of variables.

The number of variables will still determine the computational speed of estimating the model. Different structure learning algorithms will entail different trade-offs between speed and accuracy (Scutari et al., 2018). However, in general the complexity of most structure learning is quadratic in the number of variables if we assume that the network we are trying to learn is sparse. In other words, heuristic score-based algorithms with compute $O(N^2)$ network scores, and constraint-based algorithms will perform $O(N^2)$ conditional independence tests. If we do not assume the sparsity of the network, then the complexity becomes larger $O(eN)$ making it difficult to perform learning in the first place. Similar considerations pertain to parameter learning: If nodes have few parents, then local distributions will have few parameters and it is possible to estimate them efficiently.

Interestingly, the increase in the number of network scores or conditional independence tests evaluated during structure learning does not increase the likelihood of errors, as the literature on multiple error adjusting would suggest. It is widely recognized that all structure learning algorithms in this paper are empirically self-adjusting for multiplicity, even though no theoretical characterization of this phenomenon is available in the literature.

The quality of the data is also important. For instance, it is possible to learn BNs from incomplete data (that is, not all values are observed for all variables in all observations) by incorporating expectation-maximization and other classic statistical techniques in structure and parameter learning (Scutari, 2020). It is also possible to learn BNs from the heterogeneous data that arise from trials in which observations are collected under different conditions or with different protocols (Azzimonti et al., 2020). However, handling such issues requires additional assumptions on how the data deviate from the aforementioned assumptions and it reduces power in both structure and parameter learning.

Bayesian Networks or Pairwise Markov Random Fields?

Empirical researchers may ask whether it is better to perform a network analysis using BNs or pairwise Markov random fields, and how to compare the results from each approach. The difference between the two models from the example data set is shown in Figure 3.

Although pairwise Markov random fields allow for studying the interplay among different symptoms, and identify the most interconnected symptoms (Haslbeck & Fried, 2017), they cannot distinguish whether the interconnected symptoms are more likely to cause or be caused by surrounding variables: BNs can identify which causal

relationships are admissible in the data set, and therefore complement the information given by current network inference measures estimated from pairwise Markov random fields, given the several critical assumptions, in particular the acyclicity and the absence of latent variables.

The same applies to colliders ($A \rightarrow C \leftarrow B$), which are important structures albeit scarcely discussed in the network literature: In pairwise Markov random fields, underlying collider structures can introduce spurious negative edges among variables (De Ron et al., 2019). This means researchers may conclude from a pairwise Markov random field that two symptoms A and B are negatively connected (that is, when Symptom A is strong, Symptom B is more likely to be weak, and vice versa) when in reality a BN may show an underlying collider structure that has A and B as unconnected parents of a third node C , if a BN is the true underlying structure.

Finally, researchers should choose the model (pairwise or directed) that fits best their study goal and design (Borsboom et al., 2021): For instance, researchers may prefer BNs for studying the effect of intervention as nodes in a network (Blanken et al., 2019; Kossakowski et al., 2019; Mooij et al., 2020).

Assessing Goodness of Fit

It is common practice to assess how well a statistical model fits the data, as well as how well it predicts new data, to establish its practical relevance in the literature. This is more complicated for BNs than it is for classic statistical models for two reasons: their use of DAGs to model a multivariate distribution and the lack of a single variable of interest in learning.

First, it is important to note that a network score assigned to specific DAG is not itself useful for assessing the goodness of fit of a single model. It is not defined on a normalized scale: It becomes larger and larger in magnitude as the sample size increases and can become smaller and smaller in magnitude the more values the variables can take. Network scores are only useful to compare different DAGs, which is their task in structure learning. Reporting network scores relative to a predetermined reference network can only be useful when such a network can be established in a rigorous way, which is not usually possible without substantial expert knowledge.

Bootstrap resampling and model averaging can reliably assess whether the edges in the BN are reliably supported by the data, rather than the result of statistical noise or measurement error. At a higher level, we can also check that most edges make sense to experts in the substantive domain (for example, PTSD), or blacklisting clinically implausible edges and compare the scores of BNs with and without such expert knowledge (McNally, 2016).

```
BST <- boot.strength(data, R = 200,
  algorithm = "pc.stable",
  debug = TRUE,
  cpdag = TRUE)
# compute the edge strengths
avgnet1 <- averaged.network(BST,
```

```
  threshold = .85)
# compute the average network
astr1 <- arc.strength(avgnet1,
  data,
  "bic-g")
# compute edge strengths
```

Checking the parameters of the BN is more straightforward, in the sense that each $Pr(X_i | \Pi_{X_i})$ is either a conditional probability table in discrete BNs (Spirtes & Meek, 1995), a linear model in Gaussian BNs (Grzegorzczuk, 2010) or a collection of linear models in conditional Gaussian BNs (Cowell, 2005), which have been all studied in detail in the statistics literature (Koller & Friedman, 2009; Maathuis et al., 2018; Neapolitan, 2004; Pearl, 2009; Scutari & Denis, 2021). There are substantial practical guidelines on how to assess goodness of fit in such cases. In a Gaussian BN, each parent is associated with a regression coefficient that expresses how its effect on the child increases or decreases for a unit change in that parent. Furthermore, the standard error associated with the node describes its noisiness in terms of the amount of variance that is not explained by the parents. In a discrete BN, conditional probabilities are assigned to each value that the node takes given the configuration of the values of its parents. In a conditional Gaussian BN, discrete nodes have the same parameters as the nodes of the discrete BN and continuous nodes have the same parameters as the nodes in a Gaussian BN, possibly in a mixture: Hence their interpretation is analogous. This does not, unfortunately, apply to the entire BN.

Finally, we can use BNs as expert systems and run queries on various combinations of nodes and check whether their answers make sense to domain experts. In this context, a query consists in computing the probability of some event involving some variable of interest taking place (for instance, a given score on a symptom X) given some evidence on other variables (for instance, given scores on other symptoms Y and Z). If experts can quantify their confidence in observing events under a set of specific circumstances, based on their domain knowledge, we can measure the discrepancy between their confidence and the BN's confidence as a measure of goodness-of-fit. This approach rests, of course, on the availability of reliable expert knowledge on the phenomenon being modeled because we are treating it as a surrogate of ground truth.

Conclusion

This work contains an introduction to BNs as well as a tutorial on estimating them in R using the three main families of structure-learning algorithms: constraint-based, score-based, and hybrid algorithms. The network obtained contains directed edges, which can be interpreted as admissible causal relationships among symptoms of mental disorders under several assumptions. BNs are models that can complement the popular partial correlation networks in the broad framework proposed by the network theory of mental disorders (Borsboom, 2017).

The field of network analysis is new and uniform methods for investigating DAGs remain unestablished. Our tutorial aims to

offer an introduction to the topic as well as accessible methods to conduct these analyses.

References

- Andersen, H. (2013). When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, 80(5), 672–683. <https://doi.org/10.1086/673937>
- Azzimonti, L., Corani, G., & Scutari, M. (2020). Structure learning from related data sets with a hierarchical Bayesian score. *International conference on probabilistic graphical models* (pp. 5–16). PMLR.
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, 2(3), 47–53. <https://doi.org/10.2307/3002000>
- Bernstein, E. E., Heeren, A., & McNally, R. J. (2017). Unpacking rumination and executive control: A network perspective. *Clinical Psychological Science*, 5(5), 816–826. <https://doi.org/10.1177/2167702617702717>
- Bird, J. C., Evans, R., Waite, F., Loe, B. S., & Freeman, D. (2019). Adolescent paranoia: Prevalence, structure, and causal mechanisms. *Schizophrenia Bulletin*, 45(5), 1134–1142. <https://doi.org/10.1093/schbul/sby180>
- Blanken, T. F., Van Der Zweerde, T., Van Straten, A., Van Someren, E. J. W., Borsboom, D., & Lancee, J. (2019). Introducing network intervention analysis to investigate sequential, symptom-specific treatment effects: A demonstration in co-occurring insomnia and depression. *Psychotherapy and Psychosomatics*, 88(1), 52–54. <https://doi.org/10.1159/000495045>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4), 175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9(1), 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- 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), 1–18. <https://doi.org/10.1038/s43586-021-00055-w>
- Briganti, G., & Linkowski, P. (2019). Item and domain network structures of the Resilience Scale for Adults in 675 university students. *Epidemiology and Psychiatric Sciences*, 29, e33. <https://doi.org/10.1017/S2045796019000222>
- Briganti, G., Kornreich, C., & Linkowski, P. (2021). A network structure of manic symptoms. *Brain and Behavior*, 11, e02010. <https://doi.org/10.1002/brb3.2010>
- Briganti, G. (2021, August 4). *A tutorial on Bayesian networks*. <https://doi.org/10.17605/OSF.IO/FNE9M/>
- Briganti, G., Fried, E. I., & Linkowski, P. (2019). Network analysis of Contingencies of Self-Worth Scale in 680 university students. *Psychiatry Research*, 272, 252–257. <https://doi.org/10.1016/j.psychres.2018.12.080>
- Briganti, G., Kempnaers, C., Braun, S., Fried, E. I., & Linkowski, P. (2018). Network analysis of empathy items from the interpersonal reactivity index in 1973 young adults. *Psychiatry Research*, 265(7), 87–92. <https://doi.org/10.1016/j.psychres.2018.03.082>
- Briganti, G., & Linkowski, P. (2020). Exploring network structure and central items of the Narcissistic Personality Inventory. *International Journal of Methods in Psychiatric Research*, 29(1), e1810. <https://doi.org/10.1002/mpr.1810>
- Briganti, G., Scutari, M., & Linkowski, P. (2020). A machine learning approach to relationships among alexithymia components. *Psychiatria Danubina*, 32(Suppl. 1), 180–187.
- Briganti, G., Scutari, M., & Linkowski, P. (2021). Network structures of symptoms from the Zung Depression Scale. *Psychological Reports*, 124(4), 1897–1911. <https://doi.org/10.1177/0033294120942116>
- Briganti, G., Williams, D. R., Mulder, J., & Linkowski, P. (2020). Bayesian network structure and predictability of autistic traits. *Psychological Reports*. Advance online publication. <https://doi.org/10.1177/0033294120978159>
- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, 8(4), e60188. <https://doi.org/10.1371/journal.pone.0060188>
- Cartwright, N., & McMullin, E. (1984). How the laws of physics lie. *American Journal of Physics*, 52(5), 474–476. <https://doi.org/10.1119/1.13641>
- Castelletti, F., Consonni, G., Della Vedova, M. L., & Peluso, S. (2018). Learning Markov equivalence classes of directed acyclic graphs: An objective Bayes approach. *Bayesian Analysis*, 13(4), 1235–1260. <https://doi.org/10.1214/18-BA1101>
- Cernis, E., Evans, R., Ehlers, A., & Freeman, D. (2021). Dissociation in relation to other mental health conditions: An exploration using network analysis. *Journal of Psychiatric Research*, 136, 460–467. <https://doi.org/10.1016/j.jpsychires.2020.08.023>
- Chickering, D. M. (2002a). Learning equivalence classes of Bayesian-network structures. *Journal of Machine Learning Research*, 2, 445–498.
- Chickering, D. M. (2002b). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(Nov), 507–554.
- Constantinou, A. C., Liu, Y., Chobtham, K., Guo, Z., & Kitson, N. K. (2021). Large-scale empirical validation of Bayesian Network structure learning algorithms with noisy data. *International Journal of Approximate Reasoning*, 131, 151–188. <https://doi.org/10.1016/j.ijar.2021.01.001>
- Cowell, R. G. (2005). Local propagation in conditional Gaussian Bayesian networks. *Journal of Machine Learning Research*, 6, 1517–1550.
- Cui, R., Groot, P., & Heskes, T. (2016). Copula PC algorithm for causal discovery from mixed data. *Joint European conference on machine learning and knowledge discovery in databases* (pp. 377–392). Springer, Cham.
- Curtiss, J., Fulford, D., Hofmann, S. G., & Gershon, A. (2019). Network dynamics of positive and negative affect in bipolar disorder. *Journal of Affective Disorders*, 249, 270–277. <https://doi.org/10.1016/j.jad.2019.02.017>
- Cussens, J. (2012). Bayesian network learning with cutting planes. In C. Fabio and P. Avi (Eds.), *Proceedings of the 27th conference on uncertainty in artificial intelligence* (pp. 153–160). AUA Press. https://eprints.whiterose.ac.uk/76731/1/p153_cussens.pdf
- Daly, R., Shen, Q., & Aitken, S. (2011). Learning Bayesian networks: Approaches and issues. *The Knowledge Engineering Review*, 26(2), 99–157. <https://doi.org/10.1017/S0269888910000251>
- Davis, L. (1991). *Handbook of genetic algorithms*. Van Nostrand Reinhold.
- Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B. Methodological*, 41(1), 1–15. <https://doi.org/10.1111/j.2517-6161.1979.tb01052.x>
- Dawid, A. P. (2010). Beware of the DAG! *Causality: Objectives and assessment* (pp. 59–86). PMLR.
- De Campos, C. P., & Ji, Q. (2011). Efficient structure learning of Bayesian networks using constraints. *Journal of Machine Learning Research*, 12, 663–689.
- De Ron, J., Fried, E. I., & Epskamp, S. (2019). Psychological networks in clinical populations: Investigating the consequences of Berkson’s bias. *Psychological Medicine*, 51, 1–9.
- Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in Medicine*, 26(18), 3385–3397. <https://doi.org/10.1002/sim.2771>
- Dupont, W. D., & Plummer, W. D., Jr. (1998). Power and sample size calculations for studies involving linear regression. *Controlled Clinical Trials*, 19(6), 589–601. [https://doi.org/10.1016/S0197-2456\(98\)00037-3](https://doi.org/10.1016/S0197-2456(98)00037-3)
- Edwards, D. (2012). *Introduction to graphical modelling*. Springer Science & Business Media.

- Elliott, H., Jones, P. J., & Schmidt, U. (2019). Central symptoms predict posttreatment outcomes and clinical impairment in anorexia nervosa: A network analysis. *Clinical Psychological Science*. Advance online publication. <https://doi.org/10.1177/2167702619865958>
- Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. <https://doi.org/10.1007/s11336-020-09697-3>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. <https://doi.org/10.1037/met000167>
- 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., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*. Advance online publication. <https://doi.org/10.18637/jss.v048.i04>
- Föllmer, H. (1980). On the global Markov property. In L. Streit (Ed.), *Quantum fields—algebras, processes* (pp. 293–302). Springer. https://doi.org/10.1007/978-3-7091-8598-8_19
- Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. <https://doi.org/10.1177/1745691617705892>
- Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., Engelhard, I., Armour, C., Nielsen, A. B. S., & Karstoft, K.-I. (2018). Replicability and generalizability of posttraumatic stress disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science*, *6*(3), 335–351. <https://doi.org/10.1177/2167702617745092>
- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are ‘good’ depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. <https://doi.org/10.1016/j.jad.2015.09.005>
- Friedman, N., & Koller, D. (2003). Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning*, *50*(1/2), 95–125. <https://doi.org/10.1023/A:1020249912095>
- Friedman, N., Nachman, I., & Peér, D. (2013). *Learning Bayesian network structure from massive datasets: The “sparse candidate” algorithm*. ArXiv. <https://arxiv.org/abs/1301.6696>
- Geiger, D., & Heckerman, D. (2002). Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *Annals of Statistics*, *30*(5), 1412–1440. <https://doi.org/10.1214/aos/1035844981>
- Geiger, D., Verma, T., & Pearl, J. (1990). d-separation: From theorems to algorithms. In M. Henrion, R. D. Shachter, L. N. Kanal and J. F. Lemmer (Eds.), *Machine intelligence and pattern recognition* (Vol. 10, pp. 139–148). Elsevier.
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. <https://doi.org/10.1177/1745691614551642>
- Glover, F. (1990). Tabu search: A tutorial. *Interfaces*, *20*(4), 74–94. <https://doi.org/10.1287/inte.20.4.74>
- Goudie, R. J., & Mukherjee, S. (2016). A Gibbs sampler for learning DAGs. *Journal of Machine Learning Research*, *17*(30), 1–39.
- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*(1), 37–48. <https://doi.org/10.1097/00001648-199901000-00008>
- Grzegorzczak, M. (2010). An introduction to Gaussian Bayesian networks. In Q. Yan (Ed.), *Systems biology in drug discovery and development* (pp. 121–147). Humana Press. https://doi.org/10.1007/978-1-60761-800-3_6
- Grzegorzczak, M., & Husmeier, D. (2008). Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Machine Learning*, *71*(2-3), 265–305. <https://doi.org/10.1007/s10994-008-5057-7>
- Haslbeck, J. M. B., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *Psychological Medicine*, *47*(16), 2767–2776. <https://doi.org/10.1017/S0033291717001258>
- Haslbeck, J. M. B., & Waldorp, L. J. (2016). *mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data*. ArXiv. <https://arxiv.org/abs/1510.06871v2>
- Heckerman, D., & Geiger, D. (1995). Learning Bayesian networks: A unification for discrete and Gaussian domains. In P. Besnard & S. Hanks (Eds.), *Proceedings of the eleventh conference on uncertainty in artificial intelligence* (pp. 274–284). Morgan Kaufmann.
- Hinze, V., Ford, T., Evans, R., Gjelsvik, B., & Crane, C. (2021). Exploring the relationship between pain and self-harm thoughts and behaviours in young people using network analysis. *Psychological Medicine*, *51*, 1–10. <https://doi.org/10.1017/S0033291721000295>
- Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus [Contribution to the theory of ferromagnetism]. *Zeitschrift Für Physik*, *31*(1), 253–258. <https://doi.org/10.1007/BF02980577>
- Isvoranu, A.-M., & Epskamp, S. (2021). *Which estimation method to choose in network Psychometrics? Deriving guidelines for applied researchers*, PsyArXiv. <https://doi.org/10.1037/met0000439>
- Jones, P. J. (2017). networktools: Assorted tools for identifying important nodes in networks (R package version 1.0.0) [Computer software]. <https://CRAN.R-project.org/package=networktools>
- Jones, P. J., Mair, P., Riemann, B. C., Mugno, B. L., & McNally, R. J. (2018). A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *53*(1), 1–8. <https://doi.org/10.1016/j.janxdis.2017.09.008>
- Kalisch, M., & Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, *8*, 613–636.
- Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., & Bühlmann, P. (2012). Causal inference using graphical models with the R Package pcalg. *Journal of Statistical Software*, *47*(11), 1–26. <https://doi.org/10.18637/jss.v047.i11>
- Koivisto, M., & Sood, K. (2004). Exact Bayesian structure discovery in Bayesian networks. *Journal of Machine Learning Research*, *5*, 549–573.
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.
- Korb, K. B., & Nicholson, A. E. (2010). *Bayesian artificial intelligence*. CRC Press. <https://doi.org/10.1201/b10391>
- Kossakowski, J. J., Gordijn, M. C. M., Riese, H., & Waldorp, L. J. (2019). Applying a dynamical systems model and network theory to major depressive disorder. *Frontiers in Psychology*, *10*, 1762. <https://doi.org/10.3389/fpsyg.2019.01762>
- Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, *6*, 34175. <https://doi.org/10.1038/srep34175>
- Kuipers, J., Moffa, G., Kuipers, E., Freeman, D., & Bebbington, P. (2019). Links between psychotic and neurotic symptoms in the general population: An analysis of longitudinal British National Survey data using Directed Acyclic Graphs. *Psychological Medicine*, *49*(3), 388–395. <https://doi.org/10.1017/S0033291718000879>
- Kuipers, J., Suter, P., & Moffa, G. (2018a). *Efficient sampling and structure learning of Bayesian networks*. ArXiv. <https://arxiv.org/abs/1803.07859>
- Kuipers, J., Suter, P., & Moffa, G. (2018b). *Efficient structure learning and sampling of Bayesian networks*. ArXiv. <http://arxiv.org/abs/1803.07859>
- Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press.
- Li, H., Cabeli, V., Sella, N., & Isambert, H. (2019). *Constraint-based causal structure learning with consistent separating sets*. <https://proceedings>

- .neurips.cc/paper/2019/hash/e6872f5bbe75073f8c7cfb93de7f6f3a-Abstract.html
- Lin, S., Uhler, C., Sturmfels, B., & Bühlmann, P. (2014). Hypersurfaces and their singularities in partial correlation testing. *Foundations of Computational Mathematics*, 14(5), 1079–1116. <https://doi.org/10.1007/s10208-014-9205-0>
- Luo, X. G., Moffa, G., & Kuipers, J. (2020). *Learning Bayesian networks from ordinal data*. ArXiv. <https://arxiv.org/abs/2010.15808>
- Maathuis, M., Drton, M., Lauritzen, S., & Wainwright, M. (2018). *Handbook of graphical models*. CRC Press. <https://doi.org/10.1201/9780429463976>
- Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, 63(2), 215–232. <https://www.jstor.org/stable/1403615>
- Margaritis, D. (2003). *Learning Bayesian network model structure from data*. Carnegie-Mellon University, Pittsburgh School of Computer Science.
- McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*, 86, 95–104. <https://doi.org/10.1016/j.brat.2016.06.006>
- McNally, R. J. (2021). Network analysis of psychopathology: Controversies and challenges. *Annual Review of Clinical Psychology*, 17, 31–53. <https://doi.org/10.1146/annurev-clinpsy-081219-092850>
- McNally, R. J., Heeren, A., & Robinaugh, D. J. (2017). A Bayesian network analysis of posttraumatic stress disorder symptoms in adults reporting childhood sexual abuse. *European Journal of Psychotraumatology*, 8 (Suppl. 3), 1341276. <https://doi.org/10.1080/20008198.2017.1341276>
- McNally, R. J., Mair, P., Mugno, B. L., & Riemann, B. C. (2017). Co-morbid obsessive-compulsive disorder and depression: A Bayesian network approach. *Psychological Medicine*, 47(7), 1204–1214. <https://doi.org/10.1017/S0033291716003287>
- McNally, R. J., Robinaugh, D. J., Deckersbach, T., Sylvia, L. G., & Nierenberg, A. A. (2021). Estimating the symptom structure of bipolar disorder: Energy dysregulation as a central symptom. *Journal of Abnormal Psychology*. Advance online publication. <https://doi.org/10.1037/abn0000715>
- McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6), 836–849. <https://doi.org/10.1177/2167702614553230>
- Moffa, G., Catone, G., Kuipers, J., Kuipers, E., Freeman, D., Marwaha, S., Lennox, B. R., Broome, M. R., & Bebbington, P. (2017). Using directed acyclic graphs in epidemiological research in psychosis: An analysis of the role of bullying in psychosis. *Schizophrenia Bulletin*, 43(6), 1273–1279. <https://doi.org/10.1093/schbul/sbx013>
- Mooij, J. M., Magliacane, S., & Claassen, T. (2020). Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, 21, 1–108.
- Musella, F. (2013). A PC algorithm variation for ordinal variables. *Computational Statistics*, 28(6), 2749–2759. <https://doi.org/10.1007/s00180-013-0426-5>
- Neapolitan, R. E. (2004). *Learning Bayesian networks*. Pearson Prentice Hall.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers Inc.
- Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511803161>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Hachette U.K.
- Pearl, J., & Verma, T. (1991). A theory of inferred causation. *Proceedings of the second international conference on principles of knowledge representation and reasoning* (pp. 441–452). Morgan Kaufmann.
- Pensar, J., Talvitie, T., Hyttinen, A., & Koivisto, M. (2020). A Bayesian approach for estimating causal effects from observational data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04), 5395–5402. <https://doi.org/10.1609/aaai.v34i04.5988>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Russell, S., & Norvig, P. (2002). *Artificial intelligence: A modern approach*. Prentice Hall.
- Scanagatta, M., de Campos, C. P., Corani, G., & Zaffalon, M. (2015). Learning Bayesian networks with thousands of variables. *Advances in Neural Information Processing Systems*, 28, 1864–1872.
- Scutari, M. (2010). Learning Bayesian networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3), 1–22. <https://doi.org/10.18637/jss.v035.i03>
- Scutari, M. (2020). Bayesian network models for incomplete and dynamic data. *Statistica Neerlandica*, 74(3), 397–419. <https://doi.org/10.1111/stan.12197>
- Scutari, M., & Denis, J.-B. (2021). *Bayesian networks: With examples in R* (2nd ed.). CRC Press. <https://doi.org/10.1201/9780429347436>
- Scutari, M., Graafland, C. E., & Gutiérrez, J. M. (2018). Who learns better Bayesian network structures: Constraint-based, score-based or hybrid algorithms? *Proceedings of the ninth international conference on probabilistic graphical models* (pp. 416–427). PMLR.
- Song, L., Kolar, M., & Xing, E. (2009). Time-varying dynamic Bayesian networks. *Advances in Neural Information Processing Systems*, 22, 1732–1740.
- Spirites, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Springer. <https://doi.org/10.1007/978-1-4612-2748-9>
- Spirites, P., & Meek, C. (1995). Learning Bayesian networks with discrete variables from data. *Proceedings of first international conference on knowledge discovery and data mining* (pp. 294–299). Morgan Kaufmann.
- Suter, P., Kuipers, J., Moffa, G., & Beerenwinkel, N. (2021). *Bayesian structure learning and sampling of Bayesian networks with the R package BiDAG*. ArXiv. <http://arxiv.org/abs/2105.00488>
- Suzuki, J. (2017). An efficient Bayesian network structure learning strategy. *New Generation Computing*, 35(1), 105–124. <https://doi.org/10.1007/s00354-016-0007-6>
- Tsagris, M., Borboudakis, G., Lagani, V., & Tsamardinos, I. (2018). Constraint-based causal discovery with mixed data. *International Journal of Data Science and Analytics*, 6(1), 19–30. <https://doi.org/10.1007/s41060-018-0097-y>
- Tsamardinos, I., Aliferis, C. F., Statnikov, A. R., & Statnikov, E. (2003). Algorithms for large scale Markov blanket discovery. *FLAIRS Conference*, 2, 376–380.
- Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1), 31–78. <https://doi.org/10.1007/s10994-006-6889-7>
- 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, 5918. <https://doi.org/10.1038/srep05918>
- 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>
- Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry*, 133(5), 429–435. <https://doi.org/10.1192/bjp.133.5.429>
- Zung, W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63–70. [https://doi.org/10.1001/archpsyc.1965.01720310065008\[Mismatch](https://doi.org/10.1001/archpsyc.1965.01720310065008[Mismatch)

(Appendix follows)

Appendix

Probabilistic Equivalency Between Patterns of Edges

Patterns of Edges Like $v_j \rightarrow v_i \rightarrow v_k$, $v_j \leftarrow v_i \leftarrow v_k$ and $v_j \leftarrow v_i \rightarrow v_k$ Lead to Equivalent Probability Distributions

The product of the probability of v_j , the probability of v_i given v_j and the probability of v_k given v_i , which describes the probability of $v_j \rightarrow v_i \rightarrow v_k$ is equal to: (a) the product of the probability of v_j , the joint probability of v_i and v_j divided by the probability of v_j , and the joint probability of v_i and v_k divided by the probability of v_i ; (b) the product of the joint probability of v_i and v_j divided by the probability of v_i , and the joint probability of v_i and v_k ; (c) the product of the probability of v_i , the probability of v_j given v_i , and the probability of v_k given v_i , which describes the probability of $v_j \leftarrow v_i \rightarrow v_k$; (d) the product of the probability of v_k , the probability of v_j given v_i , and the probability of v_i given v_k , which describes the probability of $v_j \leftarrow v_i \leftarrow v_k$.

$$\begin{aligned} \underbrace{\Pr(v_j)\Pr(v_i | v_j)\Pr(v_k | v_i)}_{v_j \rightarrow v_i \rightarrow v_k} &= Pr(v_j) \frac{\Pr(v_i, v_j)\Pr(v_i, v_k)}{\Pr(v_j)\Pr(v_i)} = \\ &= \frac{\Pr(v_i, v_j)}{\Pr(v_i)} Pr(v_i, v_k) = \underbrace{\Pr(v_i)\Pr(v_j | v_i)\Pr(v_k | v_i)}_{v_j \leftarrow v_i \rightarrow v_k} \\ &= \underbrace{\Pr(v_k)\Pr(v_j | v_i)\Pr(v_i | v_k)}_{v_j \leftarrow v_i \leftarrow v_k} \end{aligned}$$

Received February 1, 2021

Revision received November 23, 2021

Accepted December 6, 2021 ■