The Challenge of Generating Causal Hypotheses Using Network Models

Oisín Ryan[1], Laura F. Bringmann[2], and Noémi K. Schuurman[3]

[1]Department of Methodology and Statistics, Utrecht University
[2]Department Psychometrics and Statistics, University of Groningen
[3]Department of Methodology and Statistics, Tilburg University

Abstract

Statistical network models based on Pairwise Markov Random Fields (PMRFs) have become increasingly popular tools to analyze multivariate psychological data, in large part due to their perceived role in generating insights into causal relationships between pairs of variables. This means that in practice statistical network models are used in analagous way to causal discovery methods from the causal modeling literature. However, it is clear that in the applied literature, the role of statistical networks in causal discovery is poorly understood. In this paper we provide a treatment of how PMRFs in general, and specifically the *Gaussian Graphical Model* (GGM) for normally distributed variables, operate as causal discovery tools. Using Directed Acyclic Graphs (DAGs) and linear structural equation models (SEMs) as representations of causal structure, we outline the key assumptions are necessary for network-based causal discovery and characterize the equivalence classes of causal models that networks identify from data under those assumptions. This allows us to clarify a number of misconceptions which we have identified in the empirical literature around the interpretation of network models. These include: the interpretation of networks as causal skeletons, the ability of networks to identify chains of causal relationships, the role of collider variables and the bias they induce, and the use of networks for learning cyclic causal models.

Keywords: Network Approach, Gaussian Graphical Model, Causal Hypotheses, Directed Acyclic Graph (DAG), Conditional Dependence, Statistical Equivalence

1 Introduction

The network approach to psychopathology is a theoretical framework in which mental disorders are viewed as arising from direct causal interactions between symptoms (Borsboom, 2017; Borsboom & Cramer, 2013). In practice, researchers often aim study the underlying network structure by estimating *statistical network models* from cross-sectional data (Epskamp, Borsboom, & Fried, 2018; Epskamp, Waldorp, Mõttus, & Borsboom, 2018; Van Borkulo et al., 2014). These graphical models generally take the form of a *Pairwise Markov Random Field* (PMRF), of which the Gaussian Graphical Model (GGM), Ising Model, and Mixed Graphical Model are special cases (Epskamp, Waldorp, et al., 2018; Haslbeck, Waldorp, et al., 2020; Van Borkulo et al., 2014). The connections between variables in these networks represent conditional dependence relationships (*partial correlations* in the case of the GGM) between each pair of variables, controlling for all other variables in the model.

The popularity of statistical network models stems in part from their perceived role as tools to learn about causal structure, with many researchers proposing that connections in an estimated network model can be used to *generate causal hypotheses* about the underlying causal system (Borsboom & Cramer, 2013; Epskamp, van Borkulo, et al., 2018; Epskamp, Waldorp, et al., 2018; Robinaugh, Hoekstra, Toner, & Borsboom, 2020). Researchers who use networks in this way typically conceptualize the data-generating structure in terms of *directed causal relationships*, and use statistical relations in the network to make inferences about those relationships according to two simple heuristics (e.g., Boschloo, Schoevers, van Borkulo, Borsboom, & Oldehinkel, 2016; Epskamp, Waldorp, et al., 2018; Haslbeck & Waldorp, 2018; van Borkulo et al., 2015). In the causal modeling

literature, using statistical relationships to make inferences about causal relationships in this way is known as *causal discovery* (Eberhardt, 2017; Glymour, Zhang, & Spirtes, 2019; Pearl, 2009; Peters, Janzing, & Schölkopf, 2017; Spirtes et al., 2000). Numerous causal discovery tools have been developed in the past decades, primarily focused on uncovering causal structure in the form of Directed Acyclic Graphs (DAGs) from observational data (Pearl, 2009; Peters et al., 2017; Spirtes et al., 2000). The heuristics used for causal hypothesis generation are in fact based on the relationship between individual connections in a statistical network and individual causal relationships in an underlying DAG.

However, these heuristics alone give only a very limited account of how statistical network models work as causal discovery tools. First, there are a number of critical assumptions which are needed for this type of causal discovery to work in practice (Pearl, 2009; Spirtes et al., 2000) and which bear closer consideration by users of statistical network models. Second, the heuristics alone say little about how well the overall network is able to identify multivariate patterns of causal relationships, let alone the advantages and disadvantages of using networks in this way. Third, network heuristics are concerned only with the presence or absence of relationships, but say little about how the sign and size of connections in networks such as the GGM should be interpreted when the interest is in causal effects. A lack of clarity around these issues has in turn lead to a number of misconceptions among applied researchers, such as the interpretation of statistical networks as *causal skeletons*, identifying the presence but not direction of causal links (e.g., Boschloo et al., 2016; Haslbeck & Waldorp, 2018; van Borkulo et al., 2015). It is well known in the graphical modeling literature that networks should not be interpreted in this way, since some edges may also reflect the presence of a common effect or collider variable (Lauritzen, 1996; Spirtes et al., 2000; Wermuth & Lauritzen, 1983).

In this paper we aim to clarify the promise and problems associated with using statistical network models as causal discovery tools. We begin by reviewing the basics of statistical network models and how they relate to both DAG representations of causal structure and linear structural equation models familiar to social science researchers. Second, we describe how the PMRF and GGM models could be causal discovery tools, clarifying the assumptions needed for causal hypothesis generation to work in practice, and characterizing the so-called equivalence classes of causal structures which these models can in principle identify from data. In the third part of the paper we focus on clarifying a number of common misconceptions we see in the applied literature regarding the use of network models for causal discovery, including the identification of causal-chains, the interpretation of edge weights in the face of collider bias, and the utility of statistical network models for studying cyclic causal relationships. We end with a brief discussion of alternative methods for causal discovery, and alternative ways of using network models to test implications of a causal model.

2 Background

In this section we will give a brief overview of two different types of graphical models: The Pairwise Markov Random Field (PMRF), which is the basis of statistical network modeling, and the Directed Acyclic Graph (DAG), which is commonly used in the causal inference literature. We also review two commonly used statistical models associated with each: the Gaussian Graphical Model (GGM) and the linear Structural Equation Model (SEM).

2.1 The Pairwise Markov Random Field and Gaussian Graphical Model

The PMRF is a graphical model consisting of variables (nodes) and undirected connections (edges) between those variables. Edges are either present or absent, but do not have a particular value (weight) associated with them. The presence of an edge denotes a particular form of *conditional dependence* between variables: an edge $X_i - X_j$ is present if and only if those variables are dependent when we condition on (statistically control for) the set of all other variables in the network $\boldsymbol{X}_{-(i,j)}$. The absence of an edge denotes that X_i and X_j are independent conditional on all other variables. We denote this as

$$X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{-(i,j)}, \tag{1}$$

where $\perp\!\!\!\perp$ stands for independence between a pair of random variables and $\not\perp\!\!\!\perp$ stands for dependence (Dawid, 1979; Lauritzen, 1996). In Figure 1(a) we show an example of a PMRF made up of four

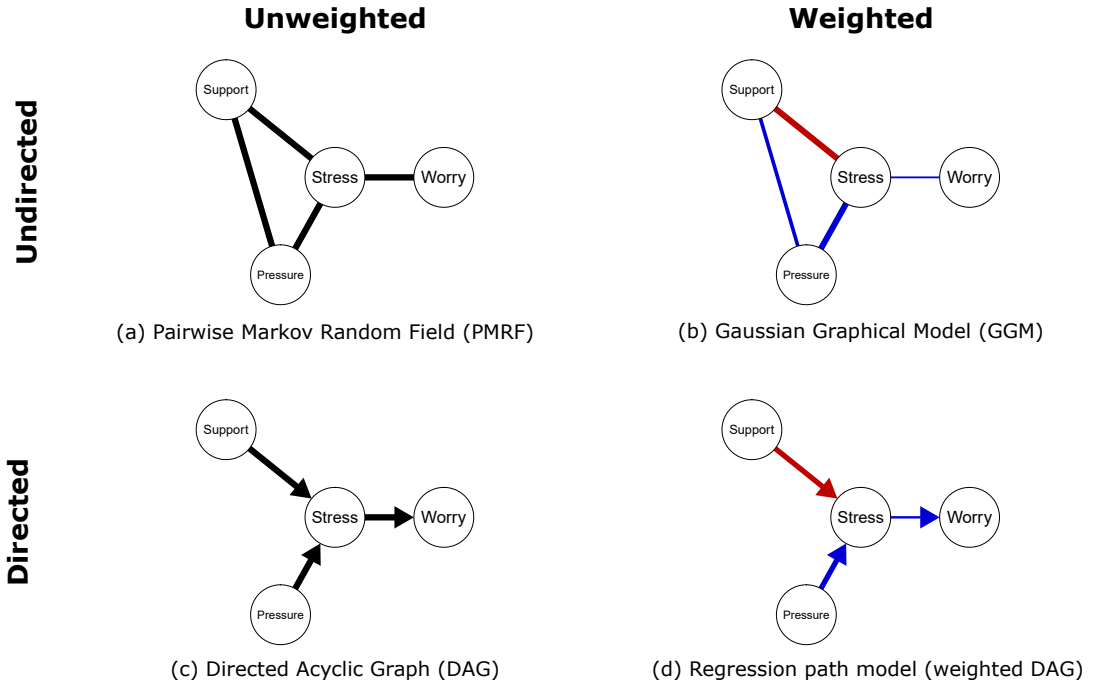

Figure 1: Four different types of network (graphical model), all describing different characteristics of the same system. In the weighted graphs, red edges depict negative relationships, blue edges depict positive edges, and the width of the edge is determined by the absolute value of the relationship (partial correlations for the GGM and regression weights for the SEM).

variables relating to burnout: levels of social support (Support), work pressure (Pressure), Stress, and Worry. From this graph, we can read off certain types of dependency relationships among those four variables. Support, Pressure and Stress are all connected to one another by undirected edges, indicating that, for instance, Pressure and Support are dependent conditional on Stress and Worry (Pressure $\not\perp\!\!\!\perp$ Support | Stress, Worry). Worry, however, is only connected to Stress: This means that Worry is dependent on Stress when we condition on Pressure and Support (Worry $\not\perp\!\!\!\perp$ Stress | Pressure, Support), but that Worry is independent of Pressure and Support conditional on Stress (Worry $\perp\!\!\!\perp$ Pressure | Stress, Support and Worry $\perp\!\!\!\perp$ Support | Stress, Pressure).

Statistical network models such as the GGM and Ising model are particular instances of PMRF models, which use assumptions about the distributions of the variables involved and their relationship with one another to produce a PMRF with *weighted* edges (Epskamp, Waldorp, et al., 2018; Van Borkulo et al., 2014). In the current paper we focus on the GGM, a popular network model for variables following a Gaussian (i.e. normal) distribution $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Cox & Wermuth, 1996; Dempster, 1972; Epskamp, Waldorp, et al., 2018; Lauritzen, 1996). The weights matrix of the GGM can be obtained by estimating the inverse of the variance-covariance matrix ($\boldsymbol{\Sigma}^{-1}$), and re-scaling this to form the matrix of *partial correlations*. These partial correlations are linear parameterizations of the conditional dependence relations defined above, and they can be interpreted in a similar way to linear regression coefficients in a model where all other variables in the network are controlled for. An example of a GGM is shown in Figure 1(b). We can interpret the positive edge connecting Stress and Worry as denoting that, keeping both Support and Pressure constant, high scores for Stress tend to co-occur with high scores for Worry, with the size of the edge denoting the strength of that relationship. Here again the the absence of an edge is taken to denote that variables are conditionally independent (i.e. have a zero partial correlation) given all other variables in the network.

2.2 Directed Acyclic Graphs and Linear SEMs

A Directed Acyclic Graph (DAG) is a graphical model consisting of unweighted directed edges between variables, with "acyclic" referring to the property that no cycles or feedback loops (e.g. $X_i \leftrightarrow X_j$) are permitted between variables. DAGs (also known as Bayesian Networks) are commonly

used to represent structural causal relationships between variables, where the directed edge $X_i \rightarrow X_j$ denotes that an intervention to change the value of X_i will lead to a change in the probability distribution of X_j (for an accessible introduction to the use of DAGs in causal inference see, amongst others Hernan & Robins, 2010; Pearl, Glymour, & Jewell, 2016; Rohrer, 2018). The DAG structure is often described in "familial" terms: Directed edges $X_i \rightarrow X_j$ connect *parent* nodes or causes (X_i) to *children* nodes or effects (X_j). Nodes which share a common child, but are not directly connected to one another, are termed *unmarried* parents, while children as well as children *of* children (and so forth) are called *descendants*. An example of a DAG is shown in Figure 1(c), where Support and Pressure are unmarried parents of Stress, Stress is a parent of Worry, and we can say that Worry is a descendant of Support, Pressure and Stress.

As a graphical model, the structure of a DAG describes which variables in $\boldsymbol{X}$ are conditionally dependent and independent from one another, but does so in a different manner to the PMRF. A DAG describes the conditional dependencies present in a set of random variables $\boldsymbol{X}$ according to the so-called *local Markov* condition (Spirtes et al., 2000). This condition states that each variable X_i is conditionally independent of its non-descendants, given its parents:

$$X_i \perp\!\!\!\perp \boldsymbol{X}_{-de(i)} \mid \boldsymbol{X}_{pa(i)}, \tag{2}$$

which can be used to read off conditional (in)dependencies between pairs of variables from the DAG. For example, in Figure 1(c), we can derive that Support and Pressure are *marginally independent*, that is, independent when not conditioning on any other variables (Support $\perp\!\!\!\perp$ Pressure $\mid \emptyset$), because Pressure is a non-descendant of Support, and Support has no parents in this graph (pa(Support) $= \emptyset$).

Further conditional (in)dependency relationships between any pair of variables can be derived from the graph using *d-separation* rules (Pearl, 2009). The most important of these rules for the current paper relates to situations in which two variables share a common effect, also known as a *collider* structure $X_i \rightarrow X_k \leftarrow X_j$. According to d-separation rules, the parent variables X_i and X_j are dependent conditional on their common child X_k. In Figure 1(c), although Support and Pressure are marginally independent, they are dependent when conditioning on Stress (Support $\not\perp\!\!\!\perp$ Pressure | Stress). This implies that if, for example, we were to statistically control for levels of stress, or for instance if we sampled only individuals with high stress levels, we would expect to see a non-zero statistical dependency between Support and Pressure.

While the DAG describes which variables are causally dependent on one another, it does not specify a form for those causal relationships or make any distributional assumptions about the variables involved. So, just as we did for the PMRF, to specify a weighted version of the DAG we need to specify the functional forms of the causal relationships between each variable, as well as the distribution of the noise term (residuals) of each variable (known as a structural causal model; Pearl, 2009). For example, we could choose to impose a Gaussian noise term for each variable, as well as linear relationships between variables, giving us a linear Structural Equation Model (SEM; Bollen, 1989).[1] The causal system would then be described by a multivariate regression model in which child nodes are predicted by their parents

$$\boldsymbol{X} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{X} + \boldsymbol{\epsilon} \tag{3}$$

where $\boldsymbol{B}$ is a $p \times p$ matrix of regression coefficients, $\boldsymbol{\alpha}$ represents a $p \times 1$ vector of intercepts, and $\boldsymbol{\epsilon}$ represents a $p \times 1$ vector of residuals. The residuals are Gaussian distributed with means equal to zero, and a diagonal variance-covariance matrix $\boldsymbol{\Psi}$, that is, the residual terms are uncorrelated. We can use the matrix of regression weights $\boldsymbol{B}$ to create a graphical representation of the SEM model as a weighted DAG, shown in Figure 1(d). Here we can see for instance that Support has a negative effect on Stress, while Pressure has a positive effect on Stress, and Stress a weaker positive effect on Worry.

There are typically many different SEM models which can be estimated from the same dataset and which will fit the data equally well, a concept known as *statistical equivalence* in the SEM literature (Bollen, 1989). This means that on the basis of statistical information alone we typically cannot distinguish between a variety of different statistically equivalent, but causally distinct, SEM models. This makes the task of discovering *the* causal structure from observational data alone typically quite difficult, even under highly idealized conditions. In the sections to follow we will see that this difficulty also extends to the use of network models for causal discovery.

[1] Note we consider only the structural equation here, omitting the measurement equation. SEM models without measurement equations are sometimes referred to as path models.

3 Network Models as Causal Discovery Tools

Many researchers use the statistical dependencies estimated by statistical network models, along with heuristics described by network methodologists, to make inferences about the underlying causal relationships between those variables. In the network literature this is sometimes referred to as *generating causal hypotheses*, a task which is equivalent to that of *causal discovery* (Eberhardt, 2017; Glymour et al., 2019; Pearl, 2009; Peters et al., 2017; Spirtes et al., 2000). In this section we examine the utility of statistical network models for discovering underlying causal DAGs. We show how PMRFs and GGMs can in principle identify aspects of causal structure, and we clarify some of the key assumptions necessary for this to work in practice. We primarily consider the case where the *population level* statistical network is known, before discussing considerations which should be additionally made when using sampled data and estimated statistical networks.

3.1 Causal Discovery with PMRFs

To be able to use a statistical network models for causal discovery, we need to know how the structure of the network model maps onto the type of causal structure we are interested in learning about. From the graphical modeling literature we know that there is a straightforward relationship between the structure of a DAG and the structure of a PMRF (Lauritzen, 1996; Spirtes et al., 2000). Specifically, when the underlying causal structure is a DAG, then the corresponding PMRF is equivalent to the *moral graph* of that DAG. The moral graph is an undirected graph obtained by first "marrying" (i.e., drawing an undirected edge between) all "unmarried parents" in the DAG, and then replacing all directed edges with undirected edges. The moral graph therefore contains an undirected edge if either a) these two nodes are connected by a directed edge in the DAG or b) these two nodes are involved in a collider structure (Lauritzen, 1996; Spirtes et al., 2000; Wermuth & Lauritzen, 1983).[2] For example, comparing Figure 1(a) and 1(c) we can see that the PMRF contains an undirected version of all of the edges in the DAG, in addition to an edge connecting the unmarried parents Support and Pressure, just as we would expect from the definition of the moral graph.

The equivalence of the PMRF and the moral graph of a DAG has allowed researchers to define two heuristics which can be used to make inferences about a DAG from a PMRF under ideal conditions (Borsboom & Cramer, 2013; Epskamp, van Borkulo, et al., 2018; Epskamp, Waldorp, et al., 2018):

> *Heuristic 1:* An edge between two variables in the PMRF ($X_i - X_j$) indicates that two variables share either a direct causal link ($X_i \rightarrow X_j$ or $X_i \leftarrow X_j$) or a common effect ($X_i \rightarrow X_k \leftarrow X_j$)
>
> *Heuristic 2:* The absence of an edge between two variables ($X_i \nsim X_j$) indicates that these two variables do not share a direct causal link ($X_i \nrightarrow X_j$ and $X_i \nleftarrow X_j$)

We can see how these heuristics work by considering what inferences could be made about the DAG structure in Figure 1(c) using only information from the PMRF in Figure 1(a). Applying the first heuristic, we would correctly identify edges in the PMRF as as indicative of either a direct causal relationship *or* the presence of a collider structure. The absence of an edge in this case is actually more informative, because it tells us that two variables are causally independent, according to the second heuristic.

A limitation of (solely) relying on these heuristics compared to other discovery methods is that these heuristics don't say anything about identifying the direction of any particular effect (Glymour et al., 2019; Peters et al., 2017; Spirtes et al., 2000). However, in principle – under certain idealizing assumptions – these heuristics are entirely valid ways to make inferences from individual statistical (in)dependencies in a PMRF to individual causal (in)dependencies in an underlying DAG. Inherent in these heuristics, however, is a critical degree of uncertainty: According to the first heuristic, we cannot use an edge in the PMRF to definitively say whether or not a directed causal relationship exists between those two variables. This is a second limitation of using the PMRF for causal discovery, and stems from the fact that the network only tests for a specific type

[2] A collider structure $X_i \rightarrow X_k \leftarrow X_j$, which does not contain an edge directly connecting the parents ($X_i \nrightarrow X_j$ and $X_i \nleftarrow X_j$) is called an *open v-structure*. If there are any open v-structures in the DAG, then the moral graph must contain an undirected edge between the relevant parents $X_i - X_j$. Thus, the moral graph "marries" unmarried parents.

of conditional independence: independence conditional on all other variables in the network. In so doing we necessarily condition on any collider variables present in the system as well. As a result, edges can be present in the PMRF which are not indicative of a direct causal relationship.

More formally, we can characterize the PMRF for causal discovery purposes as a method of identifying the *moral-equivalence set* of the underlying DAG. By a moral-equivalence set we mean that collection of DAG structures which all produce the same moral graph, and so the same PMRF. Generally speaking, there will be many DAGs which have the same moral graph, that is, there is a one-to-many mapping from PMRF to DAG structure. In Figure 2 we show the moral-equivalence set for our four-variable burnout PMRF, derived using the discovery heuristics described above along with d-seperation rules.

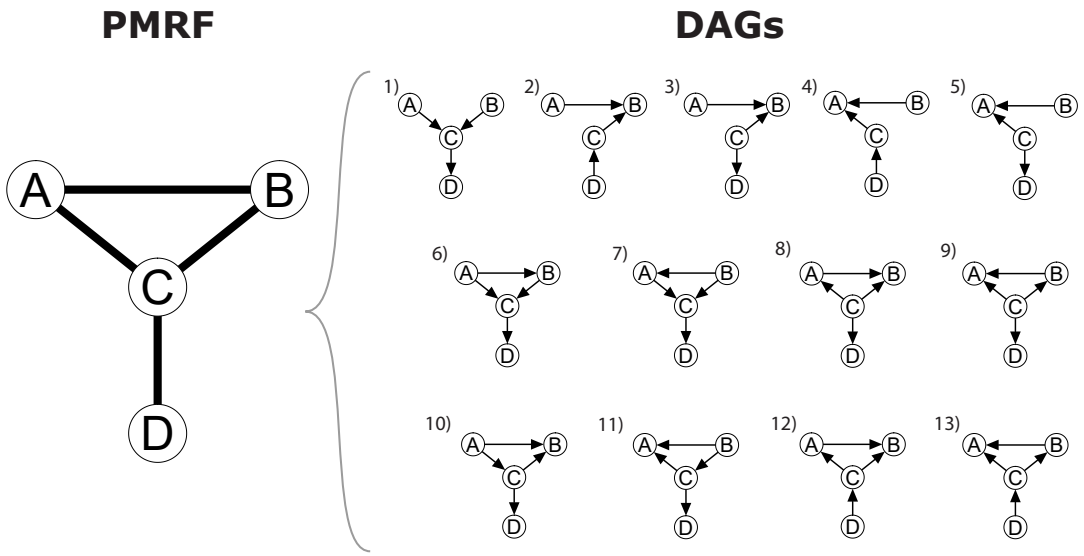

Figure 2: Visualization of the moral-equivalence set: A Pairwise Markov Random Field (PMRF) and the set of sufficient and faithful DAGs which generate that PMRF.

We can see that the moral equivalence set does contain the DAG belonging to the true causal model, but also a number of other possibilities. In fact, here 13 distinct DAG structures result in that same PMRF, and on the basis of the PMRF alone, we cannot distinguish which of these is the true underlying causal structure.

From the moral equivalence set, we can see that the heuristics described above give a correct but limited view of how PMRFs work as a DAG discovery method. We can gain more information about the possible underlying causal structure by taking into account the structure of the PMRF as a whole, rather than by inspecting only individual edges. For instance, *Heuristic 1* tells us that an edge in the PMRF might indicate a direct causal relationship, *or* the presence of a collider structure. Indeed from Figure 2 we can see that there is at least one DAG in which either A and B, A and C, or B and C are causally independent of one another, and so are only connected in the PMRF due to the presence of a collider structure. However, by deriving the moral equivalence set, we can see that this is never the case for the relationship between C and D: there is always a direct connection between these two variables. Since D is only connected to C in the PMRF, it cannot be the case that C and D share a common effect. In this instance, because of the overall structure of the PMRF, and the relationship between PMRF and DAG, we can say that C and D must share a direct effect.

These types of deductions follow exactly the same logic as the heuristics described above, but follow from considering the graph as a whole, using the moral-equivalence set, rather than focusing only on individual edges in the network: By considering what *combinations* of relationships are and are not allowed by the PMRF we could gain more insight into the underlying causal structure. Using only the hypothesis-generation heuristics above, it can be difficult to read off these global implications of the PMRF, but as a causal discovery method for DAGs, this is the extent of the information the PMRF gives us. In the following, we discuss the critical assumptions which are needed to make inferences about causal relationships based on PMRFs – using either the heuristics

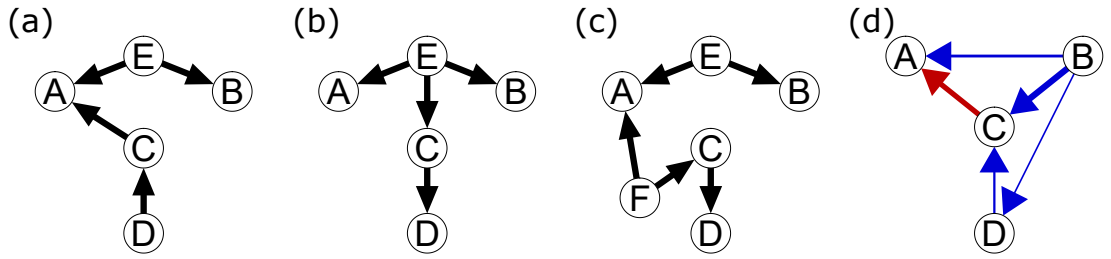

Figure 3: Four DAGs which generate the same PMRF between the variables A, B, C and D, as shown in Figure 2, but which violate some assumption which is made in deriving the moral-equivalence set. The DAGs in panels (a), (b) and (c) represent sufficiency violations, with unobserved common cause(s) E (and F). The linear SEM in panel (d) violates faithfulness, as the directed positive relationship between B and D is exactly canceled out by conditioning on C.

described in the network literature or the moral-equivalence set – and which we have left implicit so far.

3.2 Assumptions Necessary for Causal Identification

The validity of using PMRFs for causal discovery, even when the population PMRF is known, is contingent on at least two idealizing assumptions. These assumptions are well known in the causal discovery literature, but their practical implications are not always easy to oversee.

The first of these assumptions is known as *sufficiency*: The assumption that there are no unobserved common cause variables that would induce a statistical relationship between two variables, while there is actually no causal relationship between them (Lauritzen, 1996; Pearl, 2009; Spirtes et al., 2000).[3] In general, without the sufficiency assumption, the causal information conveyed by the *presence* of an edge (as stated in *Heuristic 1*) becomes less certain, as there are typically many potential unobserved variables which may explain a particular conditional dependency between two variables. Relaxing the sufficiency assumption means that the task of causal discovery using network models becomes much more difficult, as, depending on the number and type of missing variables we are willing to consider, there are many more DAGs which may have generated the given PMRF.

For example, we may consider that a single unobserved variable, E, could act as a common cause of A and B. Then, the DAG in panel (a) of Figure 2 needs to be considered as a plausible alternative underlying causal structure in addition to those in Figure 1: This DAG also produces the same conditional dependencies present between the *observed variables* A, B, C and D as represented in the PMRF in Figure 2. In Appendix A we show that nine additional causal structures must be considered if we allow only this exact type of unobserved common cause. If we relax the sufficiency assumption even further, for example allowing E to be a common cause of more than two variables, as in panel (b) of Figure 3 or allowing more than one unobserved common cause as in panel (c) of Figure 3, deriving a complete set of possible underlying DAGs quickly becomes infeasible. In this way, it is critical to evaluate the validity of the sufficiency assumption whenever we use conditional dependence information to infer the presence of a direct causal relationship between variables.

The second major assumption needed to generate causal hypotheses using PMRFs is known as *faithfulness* (Pearl, 2009; Spirtes et al., 2000). Faithfulness is concerned with our ability to make inferences about the *absence* of a causal relationship between two variables in the face of the absence of a (conditional) statistical relationship between those two variables. To be precise, a DAG and associated probability distribution P meet the faithfulness condition if and only if every conditional (in)dependence relation in P is entailed by the local Markov condition, as described in Equation 2 (Spirtes et al., 2000). Faithfulness implies that if, say, two variables X_i and X_j are marginally independent, then by faithfulness the corresponding DAG should have no directed paths which can be traced from X_i to X_j, e.g. $X_i \to X_k \to X_j$. This means that we assume away the possibility that X_i and X_j are connected by two different directed pathways, which when combined in the marginal relationship between X_i and X_j, *exactly cancel one another out*. For example,

[3] More precisely we could say that there are no unblocked back-door paths, based on d-separation rules, passing through unobserved variables, that connect any pair of observed variables (Pearl, 2009).

faithfulness would be violated if there was a negative direct pathway $X_i \rightarrow X_j$ as well as a positive indirect pathway of equal size through X_k.

In general, without the faithfulness condition, the causal information conveyed by the absence of an edge (as in *Heuristic 2*) becomes less certain. In panel (d) of Figure 3 we show a linear SEM which would result in a violation of faithfulness. In this causal system, both B and D have a positive direct effect on C, making it a collider between them. Conditioning on this collider induces a *negative* conditional relationship between B and D. However, simultaneously, B has a *positive* direct effect on D: When combined, this negative and positive relationship cancel one another out, and so B and D appear to be conditionally independent given C - the partial correlation of B and D given C is zero, so they are unconnected in the PMRF (see Appendix A for details). In this way, it is crucial to evaluate the validity of the faithfulness assumption whenever we use conditional independence information to make inferences about causal independence between two variables (Spirtes, Meek, & Richardson, 1995).

Of course, these assumptions, and the challenge of one-to-many mapping from statistical to causal structure, are shared across many different approaches to causal discovery, although the specific details, limitations and capabilities of other approaches differ from those of the PMRF as a discovery tool. For instance, one may consider the assumption of sufficiency to be unreasonable in many psychological settings (Rohrer, 2018). Although other causal discovery methods have been adapted to deal with violations of sufficiency (Spirtes et al., 2000; Strobl, 2019; Zhang, 2008), it is unclear how the PMRF could be adapted in such a way, beyond the statement that network edges may only reflect the presence of unobserved common causes. As such, we should consider the discovery heuristics described above as describing what inferences can be made in a highly idealized setting. In the following we review the practice of using weighted PMRF models, specifically the GGM, as a causal discovery tool, and the assumptions necessary for this approach.

3.3 Causal Discovery with GGMs

The heuristics used by network researchers describe only what types of inferences can be made regarding the presence or absence of a causal relationship based on the presence or absence of a statistical relationship. Intuitively though, it would seem that more information about the underlying causal structure could be gleaned by using weighted networks such as the GGM and Ising model. These models provide additional information by parameterizing the conditional dependence relationships between variables, and so allow us to consider the sign and strength of different relationships, not only their presence or absence. The use of these models means applying additional assumptions about the functional form of the conditional dependence relationships between variables, and typically about the joint distribution of the variables involved. For instance, if the true causal relationship between two variables is non-linear, then the partial correlations used by the GGM may fail to detect, or yield a biased estimate of, that causal relationship.

To evaluate how the GGM works as causal discovery tool, let's assume that the statistical assumptions of the GGM are met, that is, the variables involved have a joint Gaussian distribution, and the causal relationships between them are linear in form, as in Equation 3. This means we can view the true underlying causal structure as a linear SEM. From the SEM literature, we know that there is a straightforward expression relating the parameters of such a SEM to a (model-implied) variance-covariance matrix:

$$\boldsymbol{\Sigma} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{B})^{-T} \tag{4}$$

where $\boldsymbol{\Psi}$ is the residual variance-covariance matrix of $\boldsymbol{X}$, diagonal in the case of uncorrelated residuals (Bollen, 1989). Re-calling that the weights matrix of the GGM is simply the inverse of the variance-covariance matrix ($\boldsymbol{\Sigma}^{-1}$, this expression allows us to understand how the GGM maps onto the linear SEM model, and so, how the network works as a causal discovery method under the most ideal conditions possible.

The relationship between the GGM and the linear SEM model has many of the same implications for causal discovery as those of the PMRF and DAG relationship we reviewed above (Epskamp, Waldorp, et al., 2018). First, if we are willing to assume sufficiency, then the presence of an edge in the GGM will correspond to either the presence of a weighted edge in the SEM model, or the presence of a collider structure. Second, the absence of an edge in the GGM may in many situations correspond to the absence of an edge in the true path model, but this is not necessarily always the case, as we discuss below. The most important implication of this relationship is that there is a one-to-many mapping from GGM weights matrix to linear SEM model. As we noted earlier,

there are typically many different SEM models which are statistically equivalent to one another, a notion which we can understand here to mean that there are many different sets of regression parameters $\boldsymbol{B}$ that can produce the same variance-covariance matrix $\boldsymbol{\Sigma}$ (MacCallum, Wegener, Uchino, & Fabrigar, 1993; Raykov & Marcoulides, 2001). By implication, even though there is a one-to-one mapping from variance-covariance matrix to GGM, there is a one-to-many mapping from a GGM to an underlying linear causal model.

We can understand the utility of the GGM as a causal discovery tool by deriving this set of statistically equivalent models from the GGM weights matrix, much as we did in the previous section for the PMRF. Although this process is a little more involved than for the moral-equivalence class in the previous section, there is a relatively simple algorithm which allows researchers to do this, described in Appendix B and available in an R package *SEset* from the github page of the first author.[4] In Figure 4 we show our example burnout GGM along with the set of linear SEM models which exactly reproduce that GGM, in other words, the *statistical-equivalence set*. To obtain this set we make the assumption of *sufficiency*, and so only consider path models which consist of the same set of variables in the GGM. Note again that the statistical equivalence set does contain the true data-generating SEM model, but that this is only one of many possibilities, and cannot be distinguished from other qualitatively different SEM models based on information in the GGM alone.

Using information about the sign and strength of the partial correlations, in combination with the linearity and normality assumptions outlined above, allows us to gain much more information about the underlying causal structure than when considering only the presence or absence of conditional dependence. Comparing the statistical equivalence set to the moral equivalence set in Figure 2, we can see that the weighting information allows us to rule out the unweighted DAGs numbered two through seven in Figure 2. For example, the moral equivalence set considers a DAG in which A and C are causally independent, but share a collider B ($A \rightarrow B \leftarrow C$) to be a valid possibility, based only the presence of conditional dependence between these three variables in the PMRF. Utilizing the above assumptions, the statistical-equivalence set rules out this possibility: It isn't possible to define a linear SEM model where A and C are causally independent, but which still produces the exact partial correlations in the given GGM.

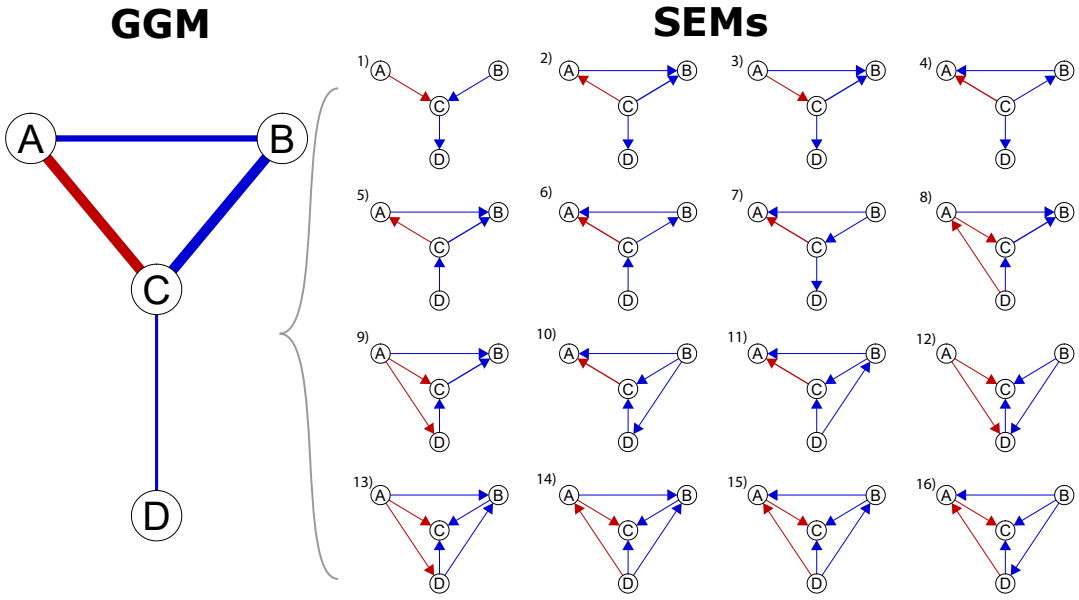

Figure 4: A Gaussian Graphical Model (GGM) and the set of DAG structure which generate that GGM.

As well as ruling out certain possibilities, the statistical equivalence set contains weighted causal structures (SEMs number eight through sixteen) which do not correspond to any DAG which appears in the moral equivalence set. For example, SEM number ten in Figure 4 contains a small

[4] https://github.com/ryanoisin/SEset

positive direct effect $B \rightarrow D$ even though the partial correlation of B and D, conditioning on both A and C, is zero in the GGM. The reason for this discrepancy is that, unlike the moral equivalence set, it is not *necessary* to invoke the assumption of faithfulness to derive the statistical equivalence set. This means that *Heuristic 2* for PMRFs is not necessarily valid for GGMs, unless the assumption of faithfulness is additionally applied. In this case, since we use the partial correlation information directly, we are able to derive different possible linear SEMs in which conditioning results in (approximately) zero partial correlations between variables.

The partial correlations in the GGM aid causal discovery by supplying the information necessary to derive the statistical equivalence set. However, it is important to note that, outside of the statistical-equivalence set, partial correlations cannot be used to make straightforward inferences about the *size* of the underlying causal effects: The mapping between linear SEM and GGM model does not guarantee that, say, the largest partial correlation also reflects the largest causal relationship. In Figure 4 the partial correlations $A - C$ and $B - C$ are the strongest edges in the GGM, being of opposite sign but having the same absolute value. However, the relative size of the corresponding direct effects in the underlying causal model varies depending on the structure of the underlying causal model. In only six of the SEM models in Figure 4 are the corresponding direct effects equal in size (models one and twelve through sixteen); in five of these models the direct effect linking B and C is larger (models two, three, five, eight and nine) and in the remaining five the direct effect between A and C is largest. (four, six, seven, ten, eleven).[5] As such, inferences about the size of causal effects based directly on the size of partial correlations in the GGM should, in general, be avoided.

As was the case for the PMRF, to use the GGM to generate causal hypotheses, it is first necessary to consider what type of causal structure we are trying to learn about, and what assumptions we are willing to make about features of that causal structure. Here we showed that by considering the GGM as a causal discovery tool for a linear and acyclic SEM model, and evaluating which of the necessary assumptions we find acceptable in a given situation, we can gain insights into the causal system through the statistical-equivalence set. As is the case for both the statistical and moral equivalence set, as the number of nodes and/or connections in the graph grow, generally speaking so too does the size of the equivalence-set. The implication of this is that inferences about the global structure of the directed graph become more and more difficult, since more and more equivalent causal structures can lead to the same statistical network, even under ideal conditions.

3.4 Estimated Network Models and Causal Discovery

So far, we have only considered the principles of using statistical network models for causal discovery using the relationship between the *population* network model and the underlying causal structure. Of course, in practice, researchers also must contend with estimating a particular network from data. This necessarily complicates the task of using statistical networks for causal discovery in practice, since researchers must contend with a number of additional concerns above and beyond those outlined above. For instance, in order to estimate the PMRF structure, researchers will typically have to choose some method to test for conditional independence between variables. To make inferences about causal (in)dependence between variables, we must assume that we have used an appropriate technique to evaluate their conditional independence. For instance if partial correlations are used to test for conditional independence, as in the GGM, then the functional relationship between variables should be linear. If the true causal relationships are non-linear, then we may make erroneous conclusions regarding the presence or absence of causal relationships in the true system.

When dealing with estimated network models researchers must also contend with uncertainty in their parameter estimates, that is, uncertainty about how closely the estimated network model parameters reflect the population network. Parameter uncertainty could be accounted for in many ways, for example by considering standard errors, bootstrapping the network, or by sampling different network models from the posterior in a Bayesian approach (Epskamp, Borsboom, & Fried, 2018; Williams & Mulder, 2020a, 2020b). Such an approach will almost surely lead us to considering an even larger set of possible data-generating structures than when the population network is available.

The role of parameter uncertainty is also crucial when considering the assumption of faithfulness.

[5]The example GGM and all of the SEM model weights matrices in the statistical equivalence set can be found in the supplementary materials of this paper: https://osf.io/qfyx9/

Faithfulness is an assumption regarding the presence of exact zero dependencies in the *population*. So, for example, if the GGM in Figure 4 is the population network structure, then we may consider ruling out SEMs eight through sixteen as plausible data-generating structures, since these would constitute a violation of faithfulness. However, if the GGM is an estimate of the population network structure, then we may not be so quick to rule out these possibilities. When estimating a network model, researchers typically use either a decision rule (such as a threshold value or significance test) to set certain edge weight estimates to zero, or use regularization techniques which induce bias in the parameter estimates such that small edge weights are set exactly to zero (Epskamp & Fried, 2018; Friedman, Hastie, & Tibshirani, 2008; Williams & Rast, 2018; Williams, Rhemtulla, Wysocki, & Rast, 2019). The upshot is that, even if we are willing to assume faithfulness with respect to the population, the absence of an edge in an estimated network model may not reflect the absence of an underlying causal relationship (Tzelgov & Henik, 1991; Uhler, Raskutti, Bühlmann, & Yu, 2013). This represents a challenge to the practical application of network models for causal discovery.

3.5 Intermediate Summary

Statistical network models can be understood as causal discovery tools which allow for inferences about particular *equivalence classes* of directed causal structures. The PMRF can be understood as identifying the moral equivalence class of an underlying DAG, as long as the assumptions of sufficiency and faithfulness hold. This characterization of the PMRF is consistent with the heuristics for causal hypothesis generation proposed by network methodologists. If we are willing to additionally assume that the causal structure consists of linear relationships between variables with a joint Gaussian distribution, and that sufficiency holds, then we can understand the GGM as identifying the statistical-equivalence class of an underlying linear SEM model. In general, the utility of statistical network models for causal discovery, like any causal discovery method, is dependent on the validity of assumptions we are willing to make about the underlying causal structure. The more specific assumptions we can make, and the more specific knowledge we can offer about the causal structure, the more specific the inferences we can make. Alternative methods for causal discovery exist, however, and may be more appropriate than statistical network models in the context described above. We return to this issue in the discussion.

4 Implications and Common Misconceptions

It is clear that many empirical researchers are utilizing statistical network models, either implicitly or explicitly, as causal discovery tools. However, it is also clear that there is a lack of clarity around exactly how this should work in practice. While in the previous section we have outlined the properties of network models as causal discovery tools, there are a number of implications of this treatment which may not be immediately obvious, but which have direct implications for how these models are used in empirical research.

In the following we tackle four different misconceptions that we have identified in the applied literature around statistical network models, which we believe have arisen due to a lack of clarity around how these models can be used for causal discovery. These misconceptions concern the interpretation of networks as causal skeletons, the use of network heuristics to infer multivariate patterns of causal relationships, the effect of collider variables on the size and sign of network edges, and the possibility of using network models to discover cyclic causal relationships.

4.1 I: Causal Skeletons and Spurious Edges

As outlined above, the PMRF can be interpreted as the moral graph of an underlying DAG. Crucially, this means that the moral graph of a DAG should not be confused with the *skeleton* of a DAG. The *skeleton* of a DAG is the undirected graph obtained by replacing the directed edges in a DAG with undirected edges: The skeleton contains an edge between two nodes *if and only if* the underlying DAG contains an edge between those nodes (Spirtes et al., 2000). The skeleton describes exactly which variables do and do not share a connection, but does not contain information on the directionality of that connection. In general the moral graph will contain more edges than the skeleton, with additional edges induced in the moral graph whenever there is a collider with unmarried parents. The difference between the moral graph (PMRF) and the skeleton is shown for three example DAGs in Figure 5. We see that although all three DAGs share the same

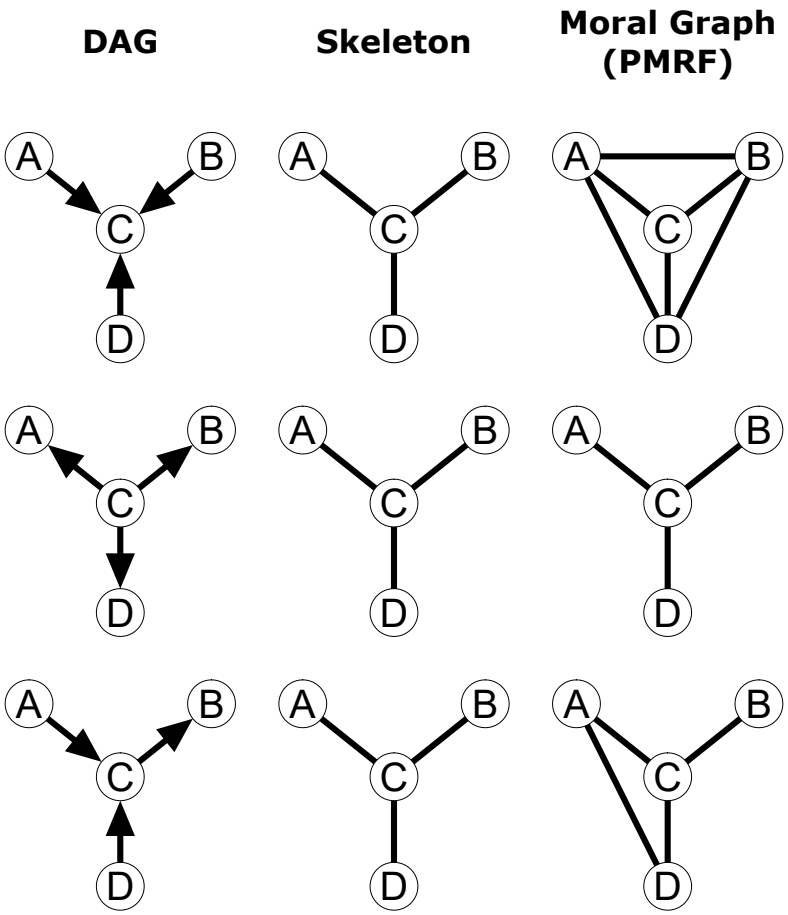

Figure 5: Three examples of DAGs which all have the same skeleton, but result in different moral graphs depending on the orientation of the edges in the DAG.

skeleton, each moral graph is distinct. If we interpret the DAG in causal terms, we can see that variables can be *conditionally dependent* in the PMRF even while being *causally independent* in the corresponding DAG: In both the first and third row of Figure 5, A and D are conditionally dependent given C (there is an edge $A - D$ in the PMRF), but intervening on A has no effect on the value of D, or vice versa.

While many researchers correctly describe the relation between PMRF and DAG (Borsboom & Cramer, 2013; Epskamp, Waldorp, et al., 2018), it is important to note here an unfortunate use of the term "causal skeleton" in relation to the PMRF (e.g Armour, Fried, Deserno, Tsai, & Pietrzak, 2017; Borsboom & Cramer, 2013; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; Haslbeck & Waldorp, 2018; Isvoranu et al., 2016; van Borkulo et al., 2015), which can easily lead to confusion regarding the causal status of edges in statistical network models. For example, researchers may be tempted to interpret that, by statistically controlling for other variables, the statistical dependencies in the network have filtered out any "spurious" relationships (Rohrer, 2018; Spector & Brannick, 2011) thereby producing a causal skeleton. The notion that networks filter out certain uninteresting relationships is often emphasized in the network literature: Costantini et al. (2015) point out that conditioning on all other variables has the effect of omitting "spurious" connections in the sense of *marginal dependencies*, as produced by mediation or common cause pathways. For instance, in the bottom row of Figure 5 the path $A \rightarrow C \rightarrow B$ implies that A and B are marginally dependent ($A \not\perp\!\!\!\perp B$), although there is no direct effect between them, and we can see that the PMRF correctly omits any $A - B$ edge. However, it is also true that by conditioning on all variables, we will typically also induce spurious relationships by conditioning on colliders, as

we can see from the $A - D$ edge in that very same network.[6]

4.2 II: Causal Chains

Although the two heuristics which are used for causal discovery with network models correctly describe how any individual edge can be interpreted, they fail to communicate information about the global structure of different possible underlying causal systems. However, it is clear from the empirical literature that very often researchers are interested in making inferences about (parts of) the global structure of the system, not just individual edges. One example of this is the common practice in empirical papers of interpreting a *chain* of dependent variables in a PMRF, of the form $X_i - X_j - X_k$ as indicative of a directed mediation structure, $X_i \rightarrow X_j \rightarrow X_k$, interpretations which are present in, for example, Deserno, Borsboom, Begeer, and Geurts (2017) (X_i = social contacts, X_j = social satisfaction, X_k = feeling happy), Isvoranu et al. (2016) (X_i = sexual abuse, X_j = anxiety and depression, X_k = psychosis) and Fried et al. (2015) (X_i = bereavement, X_j = lonely, X_k = sad and happy). These types of interpretations may be reasonable in many cases, but are prone to error if they are made without considering the appropriate equivalence set, that is, the moral- or statistical-equivalence sets we have described in the current paper. The reason for this is that combinations of different causal hypotheses based on individual edges may be *incompatible* with one another, as they may imply for example, a new collider structure, or some (in)dependence relationships which either contradict those in the PMRF or which are otherwise not present in the population.

In Figure 2 for example, we may be tempted to interpret the PMRF as showing a pattern of dependencies consistent with the hypothesis that A has a direct effect on B ($A \rightarrow B$), as well as an indirect effect through C ($A \rightarrow C \rightarrow B$). Let us call this hypothesis I. To take another example, a researcher may be tempted to hypothesize that, as there is no connection between D and B, this is indicative that the effect of D on B is fully mediated by the variable C, that is $D \rightarrow C \rightarrow B$. Let us call this hypothesis II. Inspecting Figure 2 we can see that hypothesis I holds in only one of thirteen DAG structures, DAG number 10. Furthermore, hypothesis II holds in exactly three DAG structures, numbers 2, 12 and 13. Strikingly, hypothesis I and II are totally incompatible with one another: although both hypotheses are reasonable explanations of two different local structures, and in fact hypothesize the same direction for one edge $C \rightarrow B$, it is impossible for both to be true in any of the underlying DAG structures that correspond to the given PMRF. Thus, if researchers use the hypothesis-generation heuristics to make inferences about individual edges, they may be correct under the conditions outlined above. But, if they try to combine these inferences to say something about causal chains or global structures, they may easily come to erroneous conclusions.

4.3 III: Colliders, Negative Edges and the Positive Manifold

As we have outlined above, the size and sign of edges in networks such as the GGM may not reflect in a straightforward way the value of a corresponding direct effect in the causal system being studied. Discussions of this phenomena in the applied literature, however, often revolve around a specific consequence of collider conditioning in inducing small negative edges in statistical networks (Betz, Penzel, Rosen, & Kambeitz, 2020; Black, Panayiotou, & Humphrey, 2020; Epskamp, Waldorp, et al., 2018; Hoffman et al., 2019; Isvoranu et al., 2020). In clinical psychology, the positive manifold refers to the observation that symptoms of a disorder are typically positively correlated, and one popular explanation for the positive manifold is that the underlying causal system consists only of positive (activating) relationships between nodes (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011; van Bork, Epskamp, Rhemtulla, Borsboom, & van der Maas, 2017; Van Der Maas et al., 2006). From the causal modeling literature, it is known that when two causally independent parents both exert a positive linear influence on a collider variable, (a *positive collider*), then conditioning on that collider will induce a negative statistical dependency between those variables (for details see Greenland, 2003; Jiang & Ding, 2017; Nguyen, Dafoe, & Ogburn, 2019; Pearl, 2013). An example is shown in the top row of Figure 6 where, although A and B are causally independent in the underlying linear causal model, there is a negative partial correlation between them conditional

[6] Notably, there are other reasons not considered here why relationships in a statistical network can be considered spurious or biased estimates of an (in)dependence relationship. For example, measurement error can result in under- or over-estimation of partial correlation relationships relative to those present between the latent constructs of interest (c.f., Buonaccorsi, 2010; Schuurman & Hamaker, 2019)

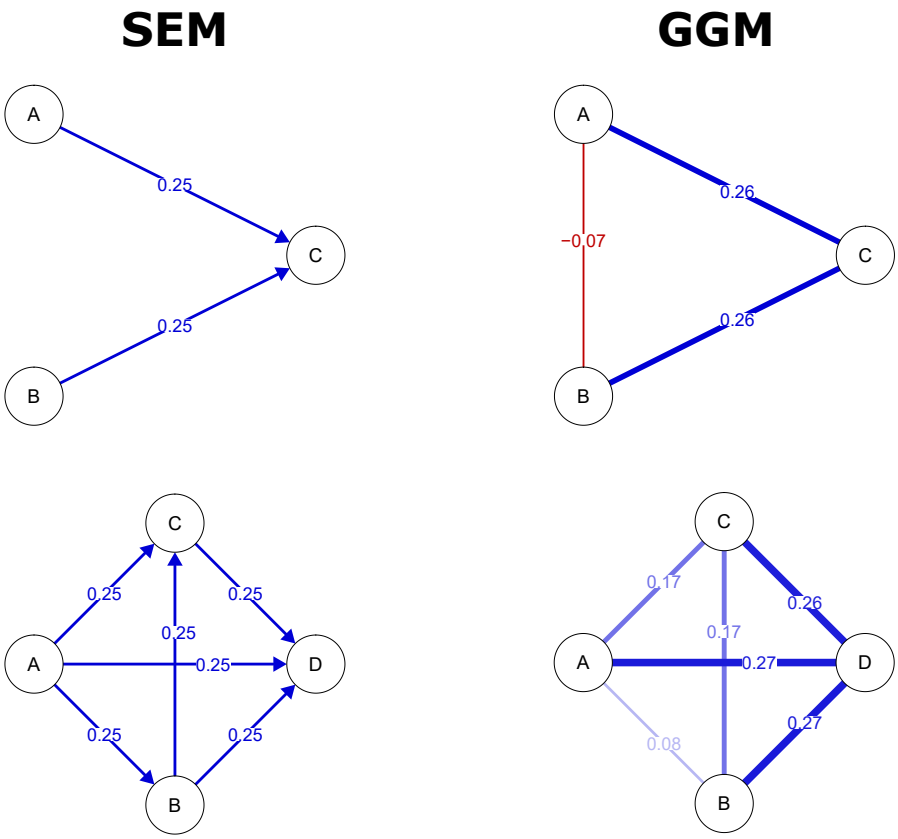

Figure 6: Two examples of linear SEMs and their corresponding GGM.

on the collider C.[7] Taken together, this means that, if we are willing to assume that there is a positive causal manifold, then small negative edges in the statistical network may be interpreted as spurious, in the sense that they are induced by collider conditioning, and not reflective of the presence of a direct causal effect (Betz et al., 2020; Black et al., 2020; Epskamp, Waldorp, et al., 2018; Hoffman et al., 2019; Isvoranu et al., 2020).

However, it would be incorrect to think that this is the *only* undesirable consequence of collider conditioning, or that the presence of colliders can always be detected in such a straightforward way even if we assume a positive manifold. A more precise formulation of collider bias is that, when we aim to estimate the direct causal dependency between a pair of variables sharing a positive collider, conditioning on that collider introduces a *negative bias* in our estimate of that causal effect. However, this negative bias does not necessarily result in a small negative edge in the network. Instead, this bias can produce, in a general sense, highly inaccurate inferences about the strength and size of causal relationships on the basis of the strength and size the statistical network relationships.

The implications of negative collider bias for network researchers are made clear if we consider a simple example. The bottom left-hand panel of Figure 6 depicts a linear SEM model reflecting a system which exhibits the positive causal manifold, and where the positive standardized regression weights are all exactly equal in magnitude. In other words, every single direct effect in the system is exactly equal in strength: increasing A by one unit has the exact same positive direct effect on B, C and D, and so forth for all other direct effects in the system. However, when we derive the corresponding GGM for this system in the bottom right-hand panel, we obtain a network with a range of different edge weights. Here, the edge between A and B is much smaller than that between

[7] Let ρ_{AC} represent the correlation between A and C and let ρ_{BC} represent the correlation between B and C. If C is a positive collider in a linear SEM model, then ρ_{AC} and ρ_{BC} lie between zero and one in value. Take it that A and B are causally independent, as in the top right panel of Figure 6. Following Pearl (2013), the partial correlation between A and B conditional on C is given by $\frac{-\rho_{AC}\rho_{BC}}{\sqrt{(1-\rho_{AC}^2)(1-\rho_{BC}^2)}}$, which, given the aforementioned range restriction, will be negative.

A and C, which is again smaller than the edge connecting A and D. The asymmetry in edge weights in the GGM is produced by the differential effects of collider bias. Since A and B share two positive colliders, their GGM edge weight experiences a greater degree of negative bias than the edge weight between A and C, while the edge weight between A and D could be considered unbiased since those variables do not share any colliders in the SEM. Researchers presented with this GGM may be tempted to interpret that A and B have only a very weak direct effect on one another, while D shares much stronger direct effects with all other variables in the networks. Perhaps researchers would even conclude that D is an attractive intervention target, since it has relatively high direct effect centrality (Bringmann et al., 2019; Robinaugh, Millner, & McNally, 2016). However, in this instance all such conclusions would be patently incorrect, since in the true causal system, all direct effects are exactly equal.

In the general case, the degree of collider bias present in any edge weight depends on the number of colliders present as well as the strength of the individual effects which make up that collider structure. This means that in the general case, without any knowledge of the direction of effects, the presence and degree of collider bias is difficult - if not impossible - to detect from the statistical network structure alone. As we saw above, this has obvious implications for researchers who attempt to interpret the magnitude of edge weights in a statistical network, or who attempt to calculate metrics like density and centrality, which are direct functions of those edge weights: Since network edge weights may be very different from their corresponding causal relationships, even when a positive manifold can be assumed, the network edge weights may produce a misleading account of the causal relationships in the system (see also Dablander & Hinne, 2019). Even if the presence of a network edge does reflect the presence of a causal effect (as in *Heuristic 1*), the weight of the network edge should not in general be taken as an estimate of that causal effect.

4.4 IV: Causal Agnosticism and Causal Cycles

Although we have focused on characterizing the utility of statistical network models as DAG discovery tools, it is also clear from the literature that many network researchers explicitly reject the idea that the underlying causal structure of interest is a DAG. In fact, many researchers adopt what could be called a *causally agnostic* approach: leaving unspecified whether the causal structure contains uni-directional or bi-directional relationships, or whether the structure is cyclic or acyclic (Costantini et al., 2015; Cramer, Waldorp, van der Maas, & Borsboom, 2010; Epskamp, Waldorp, et al., 2018; Isvoranu et al., 2016; McNally et al., 2015; Robinaugh et al., 2016; Van Borkulo et al., 2014). This state-of-affairs presents an obstacle to assessing the utility of statistical network models for causal discovery, since this agnosticism leaves it unclear how we could go about such an assessment. That is, without specifying the type of causal structure we wish to discover, it is unclear whether the heuristics used to conduct causal discovery are actually valid. To make a statistical analogy, it is impossible to evaluate the properties of a given estimator if we never specify what the target of inference (i.e. the estimand) actually is (Haslbeck, Ryan, & Dablander, 2020).

One reason why researchers reject the DAG as a model for the underlying causal structure is that the DAG by definition does not allow for the presence of *causal loops*, that is, directed cyclic causal relationships such as $A \rightarrow B \rightarrow C \rightarrow A$ or $A \leftrightarrow B$. In substantive terms these causal loops would represent feedback relationships: Feeling more stressed leads you to worry more which in turn leads you to feeling more stressed. One argument for using PMRF based methods to learn about cyclic causal models was made by Epskamp, Rhemtulla, and Borsboom (2017), who correctly point out that PMRF models can represent certain conditional dependence structures which can be produced only by cyclic, rather than acyclic graphs, as shown in the top row of Figure 7. It has been shown that d-seperation rules can be applied to cyclic as well as acyclic graphs in the linear and discrete case (Pearl & Dechter, 1996; Spirtes, 1994), which means that, in principle, the two causal discovery heuristics described above can also be used in this setting. In general however, if we are willing to consider possibly cyclic causal structures, then using PMRF models for causal discovery will typically become more difficult than if we limit ourselves to considering DAGs, since in many situations the challenges and problems with this approach, as have outlined above, will be exacerbated rather than mitigated.

When aiming to discovering cyclic causal graphs, considerations of faithfulness and sufficiency are still critical, but allowing for cycles means that the size of the equivalence class of causal models which produce the same PMRF will typically become even larger and more difficult to derive than in the acyclic case. Allowing for bi-directional causal relationships by necessity increases the number

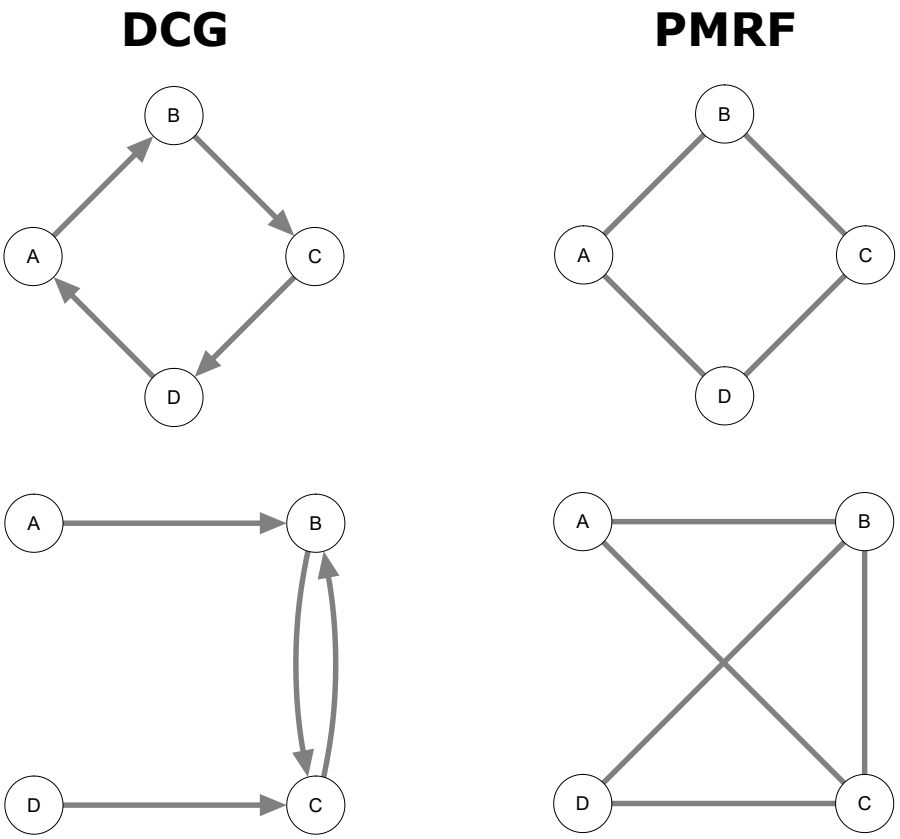

Figure 7: Two examples of Directed Cyclic Graphs (DCGs) and corresponding PMRF

of potential collider structures, as shown in the example in the bottom row of Figure 7. Here, B and C each have a direct effect on one another, meaning that we have two collider structures, $A \rightarrow B \Leftarrow C$ and $B \rightarrow C \leftarrow D$. As a result, the PMRF of this system contains two new edges induced by conditioning on a collider. Even when no new edges are induced, the possibility for more collider structures means a greater potential for collider bias impacting the edge weights in weighted networks in undesirable ways, as described above. Allowing for asymmetric pairs of direct effects between variables also raises new difficulties for the interpretation of network models: If in our example the $B \rightarrow C$ effect isn't equal to the $B \leftarrow C$ effect, how should the partial correlation between those variables in a GGM be interpreted? It's possible that statistical network models are still valuable tools to study cyclic causal systems, but researchers should be aware that the limitations and pitfalls of using these models as DAG discovery tools, as outlined in this paper, carry over to the context of using network models to discover cyclic causal relationships too.

5 Discussion

In this paper we have clarified the role of statistical network models in causal discovery, identifying and discussion a number of advantages and disadvantages of such an approach. We have fleshed out previous accounts of how network models can in principle be used for causal hypothesis generation, formalizing how network models identify particular equivalence classes of directed acyclic causal structures, and detailing the key assumptions necessary for causal discovery using statistical networks. We have also addressed a number of misconceptions which appear in the network modeling literature regarding the interpretation of networks as causal skeletons, inferences about multivariate patterns of causal relationships, the effects of collider bias and the status of edge weights as estimates of causal effects, and finally the utility of networks for learning about cyclic causal models. We believe awareness of the issues raised in this paper is critical for the growing group of researchers who wish to use statistical networks to gain insights into - that is, generate useful hypotheses about - causal structures Robinaugh et al. (2020).

As a general principle, the utility of any causal discovery method is dependent on the validity of assumptions we are willing to make about the underlying causal structure. Ironically, the more specific knowledge we have about the nature of the causal system, the less likely it becomes that the PMRF or GGM is the optimal causal discovery method to use, unless of course that system is made up of symmetric and undirected causal relationships (Cramer et al., 2016; Dalege et al., 2016). In other situations, alternative causal discovery tools which are built for purpose will likely outperform statistical network models. A number of causal discovery methods have been developed which aim to uncover causal structure using patterns of conditional independence between variables (Ramsey, Glymour, Sanchez-Romero, & Glymour, 2017; Scheines, Spirtes, Glymour, Meek, & Richardson, 1998; Spirtes et al., 2000). These methods, such as the PC algorithm (Spirtes et al., 1995), will typically yield a different and smaller equivalence class of DAGs than the PMRF (the Markov equivalence class) simply by searching for all possible marginal and conditional independence relations among pairs of variables. Many extensions of this type of approach have been developed, which can deal with violations of sufficiency and which identify different types of equivalence classes (Colombo, Maathuis, Kalisch, & Richardson, 2012; Richardson & Spirtes, 2002; Spirtes et al., 2000). Other ca8usal discovery methods leverage assumptions about the functional form of variable relations (Mooij, Peters, Janzing, Zscheischler, & Schölkopf, 2016; Peters, Janzing, & Schölkopf, 2010) or the distribution of noise terms (Shimizu, 2014; Shimizu, Hoyer, Hyvärinen, Kerminen, & Jordan, 2006) for causal discovery from observational data. Yet other approaches make use of a mix of observational and interventional data to identify parts of the underlying causal graph (Mooij, Magliacane, & Claassen, 2016; Peters, Bühlmann, & Meinshausen, 2016), an approach recently applied in the psychological literature by Kossakowski, Waldorp, and van der Maas (2021) and Waldorp, Kossakowski, and van der Maas (2021). For an accessible overview of recent causal discovery methods, readers are referred to Glymour et al. (2019).

Of course, while most causal discovery methods focus on the discovery of DAG structures, those methods may not be appropriate when we believe the underlying causal structure to consist of cycles or feedback loops between variables. As we have outlined above, the limitations of PMRF based methods for acyclic causal discovery carry over to the cyclic case, where issues such as collider bias are likely to be exacerbated. Discovery methods for cyclic causal models have also been developed (Forré & Mooij, 2018; Hyttinen, Eberhardt, & Hoyer, 2012; Lacerda, Spirtes, Ramsey, & Hoyer, 2012) but to our knowledge, their applicability in typical psychological settings has yet to be studied. The consideration of cyclic causal models in general presents new difficulties not present in the acyclic case. For example, in the SEM literature, it is well known that, in contrast to acyclic models, not all cyclic models are statistically identified from data Bollen (1989). Understanding the conditions under which cyclic causal models can be learned from psychological data also requires an understanding of how dynamic conceptualizations of causal structure map to static causal models, and how this impacts the type of data required for cyclic causal discovery. More research into the semantics of cyclic causal graphs and how they could be applied in a psychological setting is urgently needed (see Ryan & Dablander, forthcoming).

Outside of their use for causal discovery, there are many attractive reasons to use statistical network models. The Ising model, for example, has been used as a theoretical toy model, describing a causal system with fully symmetric causal relationships (Cramer et al., 2016; Dalege et al., 2016). Network models are also useful as descriptive tools, to explore and visualize multivariate statistical relationships; they allow for the identification of predictive relationships (Epskamp, Waldorp, et al., 2018); provide sparse descriptions of statistical dependency relationships in a multivariate density; and may be used as a variable clustering or latent variable identification method (e.g. Golino & Epskamp, 2017). These useful characteristics suggest that PMRFs estimated from psychological data might best be viewed as phenomena to be explained by some theoretical model of psychopathology, rather than as a tool to discover causal relationships in an inductive fashion (Haslbeck & Ryan, 2021; Haslbeck, Ryan, Robinaugh, Waldorp, & Borsboom, 2021; Robinaugh, Haslbeck, Ryan, Fried, & Waldorp, 2021). If the network approach to psychopathology is to make progress in the hunt for causal mechanisms underlying psychological disorders, further development of both theoretical modeling approaches and causal discovery methods suitable to a psychological context are in dire need. We hope that the analysis presented in the current paper helps to clarify the possible role that statistical network models might play in that hunt moving forward.

Acknowledgements

The work of Oisín Ryan was supported by a grant from the Netherlands Organisation for Scientific Research (NWO; Onderzoekstalent Grant 406-15-128). We would like to thank Denny Borsboom, Mihai Constantin, Angelique Cramer, Fabian Dablander, Sacha Epskamp, Ellen Hamaker, Jonas Haslbeck, Erik-Jan van Kesteren, IJsbrand Leertouwer, Don Robinaugh and Douglas Steinley for, variously, discussions about, feedback on and reviews of earlier drafts and presentations of this work.

References

Abadir, K. M., & Magnus, J. R. (2005). *Matrix Algebra* (Vol. 1). Cambridge University Press.

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., & Pietrzak, R. H. (2017). A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in us military veterans. *Journal of Anxiety Disorders*, *45*, 49–59.

Betz, L. T., Penzel, N., Rosen, M., & Kambeitz, J. (2020). Relationships between childhood trauma and perceived stress in the general population: a network perspective. *Psychological medicine*, 1–11.

Black, L., Panayiotou, M., & Humphrey, N. (2020). Internalizing symptoms, well-being, and correlates in adolescence: A multiverse exploration via cross-lagged panel network models. *Development and Psychopathology*, 1–15.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Oxford, England, John Wiley & Sons.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407.

Boschloo, L., Schoevers, R. A., van Borkulo, C. D., Borsboom, D., & Oldehinkel, A. J. (2016). The network structure of psychopathology in a community sample of preadolescents. *Journal of Abnormal Psychology*, *125*(4), 599.

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., ... Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892.

Buonaccorsi, J. P. (2010). *Measurement error, models, methods, and applications*. Boca Raton, FL: Chapman & Hall/CRC.

Colombo, D., Maathuis, M. H., Kalisch, M., & Richardson, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 294–321.

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29.

Cox, D. R., & Wermuth, N. (1996). *Multivariate dependencies: Models, analysis and interpretation*. Chapman and Hall/CRC.

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One*, *11*(12), e0167490.

Cramer, A. O., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2-3), 137–150. doi: 10.1017/S0140525X09991567

Dablander, F., & Hinne, M. (2019). Node centrality measures are a poor substitute for causal inference. *Scientific reports*, *9*(1), 6846.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The causal attitude network (can) model. *Psychological Review*, *123*(1), 2.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–31.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 157–175.

Deserno, M. K., Borsboom, D., Begeer, S., & Geurts, H. M. (2017). Multicausal systems ask for multicausal approaches: A network perspective on subjective well-being in individuals with autism spectrum disorder. *Autism*, *21*(8), 960–971.

Eberhardt, F. (2017). Introduction to the foundations of causal discovery. *Int. J. Data Sci. Anal.*, *3*(2), 81–91.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods*, *23*(4), 617.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network pschometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927.

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science*, *6*(3), 416–427.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480.

Forré, P., & Mooij, J. M. (2018). Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. *arXiv preprint arXiv:1807.03024*.

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O., ... Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology*, *124*(2), 256.

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.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Ghoshal, A., & Honorio, J. (2017). Learning linear structural equation models in polynomial time and sample complexity. *arXiv preprint arXiv:1707.04673*.

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, *10*, 524.

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035.

Greenland, S. (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, 300–306.

Haslbeck, J., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*, 1–32.

Haslbeck, J., Ryan, O., & Dablander, F. (2020, May). *The sum of all fears: Comparing networks based on symptom sum-scores.* PsyArXiv. Retrieved from psyarxiv.com/3nxu9 doi: 10.31234/osf.io/3nxu9

Haslbeck, J., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2021). Modeling psychopathology: From data models to formal theories.

Haslbeck, J., Waldorp, L., et al. (2020). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, *93*(8).

Haslbeck, J., & Waldorp, L. J. (2018). How well do network models predict observations? on the importance of predictability in network models. *Behavior Research Methods*, *50*(2), 853–861.

Hernan, M. A., & Robins, J. M. (2010). *Causal inference.* CRC Boca Raton, FL;.

Hoffman, M., Steinley, D., Trull, T. J., Lane, S. P., Wood, P. K., & Sher, K. J. (2019). The influence of sample selection on the structure of psychopathology symptom networks: An example with alcohol use disorder. *Journal of abnormal psychology*, *128*(5), 473.

Hyttinen, A., Eberhardt, F., & Hoyer, P. O. (2012). Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, *13*(1), 3387–3439.

Isvoranu, A.-M., Guloksuz, S., Epskamp, S., van Os, J., Borsboom, D., Investigators, G., et al. (2020). Toward incorporating genetic risk scores into symptom networks of psychosis. *Psychological medicine*, *50*(4), 636–643.

Isvoranu, A.-M., van Borkulo, C. D., Boyette, L.-L., Wigman, J. T., Vinkers, C. H., Borsboom, D., & Investigators, G. (2016). A network approach to psychosis: pathways between childhood trauma and psychotic symptoms. *Schizophrenia Bulletin*, *43*(1), 187–196.

Jiang, Z., & Ding, P. (2017). The directions of selection bias. *Statistics & Probability Letters*, *125*,

104–109.
Kossakowski, J. J., Waldorp, L. J., & van der Maas, H. L. (2021). The search for causality: A comparison of different techniques for causal inference graphs. *Psychological Methods*.
Lacerda, G., Spirtes, P. L., Ramsey, J., & Hoyer, P. O. (2012). Discovering cyclic causal models by independent components analysis. *arXiv preprint arXiv:1206.3273*.
Lauritzen, S. L. (1996). *Graphical Models* (Vol. 17). Clarendon Press.
Levina, E., Rothman, A., Zhu, J., et al. (2008). Sparse estimation of large covariance matrices via a nested lasso penalty. *The Annals of Applied Statistics*, *2*(1), 245–263.
MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*(1), 185.
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.
Mooij, J. M., Magliacane, S., & Claassen, T. (2016). Joint causal inference from multiple contexts. *arXiv preprint arXiv:1611.10351*.
Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*, *17*(1), 1103–1204.
Nguyen, T. Q., Dafoe, A., & Ogburn, E. L. (2019). The magnitude and direction of collider bias for binary variables. *Epidemiologic Methods*, *8*(1).
Pearl, J. (2009). *Causality*. Cambridge university press.
Pearl, J. (2013). Linear models: A useful "microscope" for causal analysis. *Journal of Causal Inference*, *1*(1), 155–170.
Pearl, J., & Dechter, R. (1996). Identifying independencies in causal graphs with feedback. *arXiv preprint arXiv:1302.3595*.
Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.
Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 947–1012.
Peters, J., Janzing, D., & Schölkopf, B. (2010). Identifying cause and effect on discrete data using additive noise models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 597–604).
Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.
Ramsey, J., Glymour, M., Sanchez-Romero, R., & Glymour, C. (2017). A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International journal of data science and analytics*, *3*(2), 121–129.
Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, *8*(1), 142–149.
Richardson, T., & Spirtes, P. (2002). Ancestral graph markov models. *The Annals of Statistics*, *30*(4), 962–1030.
Robinaugh, D. J., Haslbeck, J., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743.
Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366.
Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of abnormal psychology*, *125*(6), 747.
Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.
Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, *33*(1), 65–117.

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological methods*, *24*(1), 70-91.

Shimizu, S. (2014). Lingam: Non-gaussian methods for estimating causal structures. *Behaviormetrika*, *41*(1), 65–98.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, *7*(10).

Shojaie, A., & Michailidis, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika*, *97*(3), 519–538.

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*(2), 287–305.

Spirtes, P. (1994). Conditional independence in directed cyclic graphical models for feedback.

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the eleventh conference on uncertainty in artificial intelligence* (pp. 491–498).

Spirtes, P., Glymour, C. N., Scheines, R., Heckerman, D., Meek, C., Cooper, G., & Richardson, T. (2000). *Causation, Prediction, and Search*. MIT press.

Spirtes, P., Meek, C., & Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the eleventh conference on uncertainty in artificial intelligence* (pp. 499–506).

Strobl, E. V. (2019). A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *International Journal of Data Science and Analytics*, *8*(1), 33–56.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with" well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, *109*(3), 524.

Uhler, C., Raskutti, G., Bühlmann, P., & Yu, B. (2013). Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, 436–463.

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. (2017). What is the p-factor of psychopathology? some risks of general factor modeling. *Theory & Psychology*, *27*(6), 759–773.

van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, *72*(12), 1219–1226.

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.

Van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological review*, *113*(4), 842.

VanderWeele, T. J., & Robins, J. M. (2007). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *American Journal of Epidemiology*, *166*(9), 1096–1104.

Waldorp, L., Kossakowski, J., & van der Maas, H. L. (2021). Perturbation graphs, invariant prediction and causal relations in psychology. *arXiv preprint arXiv:2109.00404*.

Wermuth, N., & Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika*, 537–552.

Williams, D. R., & Mulder, J. (2020a). Bayesian hypothesis testing for gaussian graphical models: Conditional independence and order constraints. *Journal of Mathematical Psychology*, *99*, 102441.

Williams, D. R., & Mulder, J. (2020b). Bggm: Bayesian gaussian graphical models in r. *Journal of Open Source Software*, *5*(51), 2111.

Williams, D. R., & Rast, P. (2018). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*.

Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate behavioral research*, *54*(5), 719–750.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, *172*(16-17), 1873–1896.

A Moral-Equivalent DAGs: Violations of Sufficiency and Faithfulness

In this appendix we provide additional insight into moral-equivalent DAGs which violate the assumptions of either sufficiency or faithfulness.

A.1 Sufficiency-Violating DAGs

In the main text, Figure 3(a), we show a single example of a DAG which generates the PMRF depicted in the left-hand column of Figure 2, in the presence of an unobserved variable E which acts as a common cause of both A and B. If this is the data-generating DAG, we can say that variables $A - D$ do not meet the assumption of sufficiency - to derive the true DAG, the variable E is needed.

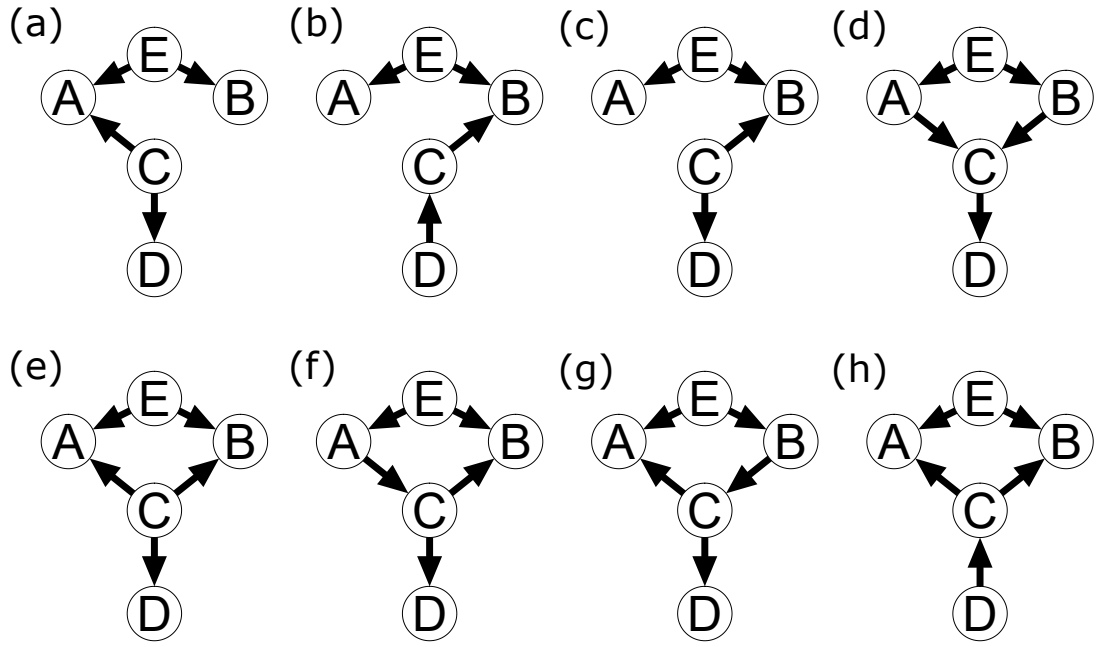

Figure 8: Eight additional DAGs which generate the same PMRF between the variables $A - D$ (as shown in Figure 2). For each DAG, the set of variables $A - D$ violates the sufficiency assumption, with respect to the variable E, a common cause of both A and B.

If we assume that sufficiency is violated only in that an unobserved variable E is a common cause of A and B, then there are eight more DAGs which generate the same PMRF for the variables $A - D$, and contain one less directed edge between the variables $A - D$ than any of the DAGs in the moral-equivalent set shown in the right-hand column. These are depicted in Figure 8. Many more moral-equivalent DAGs are possible if the assumption of sufficiency is relaxed further, that is, if different and/or additional connections between E and other variables or allowed, or if more than one unobserved common cause is allowed.

A.2 Faithfulness-Violating Weighted DAGs

In the main text, Figure 3(d), we depict a weighted DAG which generates the PMRF depicted in the left-hand column of Figure 2, but which violates the faithfulness assumption. The weights matrix of this weighted DAG is given as

$$\boldsymbol{B} = \begin{bmatrix} 0.000 & 0.200 & -0.400 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.485 & 0.000 & 0.152 \\ 0.000 & 0.100 & 0.000 & 0.000 \end{bmatrix} \tag{5}$$

Although there is a positive direct effect from B to D, denoted by the parameter $b_{41} = 0.100$, the correlation between B and D when C is conditioned on is zero. This is because conditioning on C induces a slight negative relationship between B and D, which cancels out the positive directed effect from B to D.

It is well known that if two variables X and Y are independent but share a collider Z on which both have a positive effect, there will be a negative conditional dependency between X and Y conditional on Z (cf. VanderWeele & Robins, 2007). Suppose X and Y are standard normal variables, and fully determine the variable Z, both having the same positive effect on Z. Now consider that we condition on Z taking on its average value, $Z = 0$. If we also know that X takes on a moderately low value (say, $X = -0.5$) then it must be the case that Y takes on a high value ($Y = 0.5$) - otherwise, it would not be possible for Z to take on a value of zero. In this way, although both X and Y are causally independent, if we know the value of their common effect Z, then X and Y contain information which can be used to predict one from the other.

B Statistically Equivalent SEMs from GGMs

In this appendix we describe a tool which allows us to directly relate a given GGM to a set of causal structures with linear functional relationships and Gaussian noise terms, i.e., linear SEM models. This tool takes as input an (estimated) precision matrix or partial correlation matrix and outputs a set of linear path models structures between the observed variables. While each of these path models may lead to very different substantive interpretations, they all lead to the same pattern of partial correlations between the observed variables described by the GGM. We can say that they are statistically-equivalent in the sense of a structural equation model, in that they produce the same model-implied variance-covariance matrix (Bollen, 1989; MacCallum et al., 1993; Raykov & Marcoulides, 2001; Tomarken & Waller, 2003). This tool is thus called the *statistical-equivalence-set* or *SE-set* algorithm. The current implementation is limited to deriving path models which meet the sufficiency assumption. However, since we utilize SEM matrix relationships to derive the SE-set, rather than a constraint-based causal discovery method (Pearl, 2009; Spirtes, 1995) it is not necessary to invoke the faithfulness assumption, and so the SE-set algorithm may produce DAGs which are unfaithful to the set of conditional (in)dependencies described by the GGM. For a similar method that is able to discover unfaitfhul DAG structures from data see Ghoshal and Honorio (2017). Note that the SE-set algorithm is intended primarily as an illustrative and exploratory tool.

The weights matrix of the $p-$variate GGM and the weights matrix of a weighted DAG can be related to each other through the $p \times p$ *inverse variance-covariance* matrix, known as the *precision matrix* $\hat{\boldsymbol{\Omega}}$.

$$\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}. \tag{6}$$

The weights matrix of a weighted DAG was defined in Equation (3) as the matrix of regression parameters $\boldsymbol{B}$ in which child variables are regressed on their parents. In other words, the weighted DAG can be seen as a linear SEM model with uncorrelated residual terms. From the SEM literature, there is a straightforward expression relating the parameters of such a path model to a model-implied variance covariance matrix.

$$\boldsymbol{\Sigma} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{B})^{-T} \tag{7}$$

where $\boldsymbol{\Psi}$ is a matrix of residual variances of $\boldsymbol{X}$, diagonal in the case of uncorrelated residuals (Bollen, 1989). As such, the weights matrix of both the GGM and the weighted DAG can both be seen as decompositions of some variance-covariance matrix $\boldsymbol{\Sigma}$.

We use this relationship to define the *SE-set* algorithm. If $\boldsymbol{\Psi}$ and $\boldsymbol{B}$ are both known, then they can be combined to find $\boldsymbol{\Sigma}$ using Equation (7). However, the inverse operation, solving uniquely for $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ from $\hat{\boldsymbol{\Sigma}}$ is not typically possible without additional information or assumptions. It is possible to find $\boldsymbol{B}$ directly from the covariance matrix when the *topological ordering* of the DAG is known, and the residual terms are assumed to be uncorrelated (Levina, Rothman, Zhu, et al., 2008; Shojaie & Michailidis, 2010). The topological ordering of a DAG is an ordering of nodes such that every parent comes before every child. The graph in Figure 1(c) has two valid topological orderings: {Support, Pressure, Stress, Worry} and {Pressure, Support, Stress, Worry}. If the rows and columns of the covariance matrix $\boldsymbol{\Sigma}$ are sorted according to the topological ordering, then Equation (7) gives a unique solution. In that case, $\boldsymbol{B}$ will be a lower triangular matrix with zero's on the diagonal, and Equation (7) will be equivalent to an LDL^T matrix decomposition (Abadir & Magnus, 2005).

In a system of p variables, there are $p!$ possible topological orderings. Thus, every one of the $p!$ possible orderings of the rows and columns of $\boldsymbol{\Sigma}$ produces a (possibly distinct) weights matrix $\boldsymbol{B}$. Typically the number of distinct $\boldsymbol{B}$ matrices will be less than $p!$ as some DAG structures will have more than one equivalent topological ordering.

The SE-set algorithm takes as input a $p \times p$ precision matrix and calculates a corresponding $\boldsymbol{B}$ for every $p!$ possible topological ordering of the observed variables. It does this by repeatedly re-ordering the variables in $\boldsymbol{\Sigma}$ and applying the transformation in Equation (7). This gives a set of weighted DAG models of size $p!$, the weights matrices of which are collected in the SE-set $\boldsymbol{\mathcal{B}}$. Each weighted DAG in $\boldsymbol{\mathcal{B}}$ leads to the same model-implied variance covariance matrix $\tilde{\boldsymbol{\Sigma}}$. This means that, by construction, each weighted DAG in $\boldsymbol{\mathcal{B}}$ is statistically equivalent.

After deriving all $p!$ possible weighted DAGs, the size of the SE-set $\boldsymbol{\mathcal{B}}$ can be reduced in a second step by rounding or thresholding the values in the weights matrices, and removing duplicates. In the next subsection we discuss in greater detail the mathematical details of the algorithm.