Discovering Cyclic Causal Models in Psychological Research

Kyuri Park*
Supervisor: Dr. Oisín Ryan[1]

[1]Department of Methodology and Statistics, Utrecht University

January 6, 2023

This research report is written as a *half of the thesis* format, including the introduction, background, and methods section along with the references and appendices at the end. The candidate journal for publication is *Psychological Methods*.
Keywords: cyclic causal discovery, causal inference, directed cyclic graph
Word count: 2497

1 Introduction

A fundamental task in various disciplines of science is to understand the mechanisms, that is, causal relations underlying the phenomena of interest. In psychology, for example, one of the core questions is how psychopathology comes about, with the network theory positing that mental disorder is produced by a system of direct and mutual causal interactions between symptoms (Borsboom & Cramer, 2013). In practice, empirical researchers often aim to gain insights into these causal relations by fitting statistical network models to observational data, an approach that can be characterized as a form of causal discovery (Spirtes, Glymour, Scheines, & Heckerman, 2000). However, it has been shown that the network models are likely to perform poorly as causal discovery tools; relations in the network may not reflect the direct causal effects that researchers aim to discover but can instead be produced by unwittingly conditioning on common effects or by failing to account for unobserved confounding variables (Dablander & Hinne, 2019; Ryan, Bringmann, & Schuurman, 2022).

In the field of causal discovery, using statistical (in)dependencies estimated from observational data to infer causal structure is known as *constraint-based* causal discovery (Spirtes & Glymour, 1991). Ryan et al. (2022) suggest that network models could be replaced by using purpose-built constraint-based causal discovery algorithms. However, the most popular and well-studied constraint-based algorithms assume that causal relationships are *acyclic*; if X causes Y, then Y does not cause X (Glymour, Zhang, & Spirtes, 2019). This is problematic since *cyclic* relationships or *feedback loops* are critical to the theoretical understanding of psychopathology (Borsboom, 2017).

*Correspondence concerning this paper should be addressed to: Kyuri Park, Department of Methodology and Statistics, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands. e-mail: k.park@uu.nl

For example, Wittenborn, Rahmandad, Rick, and Hosseinichimeh (2016) suggest that several different causal feedback loops, such as *perceived stress* → *negative affect* → *rumination* → *perceived stress* play a key role in sustaining depression. Such theoretical expectations necessitate the use of *cyclic causal discovery* algorithms. Although a number of cyclic causal discovery algorithms have been developed (Mooij & Claassen, 2020), they have not been as well studied as their acyclic counterparts, which is partly due to practical difficulties in fitting and interpreting cyclic causal models. To our knowledge, little to no research has been done on the applicability of constraint-based cyclic causal discovery algorithms in psychological research.

In this paper, we investigate the suitability of cyclic causal discovery methods for psychological research. The goal of this study is twofold. First, we will provide an accessible overview of different cyclic causal discovery methods. Second, we will investigate how well each of these methods works by means of a simulation study.

2 Background

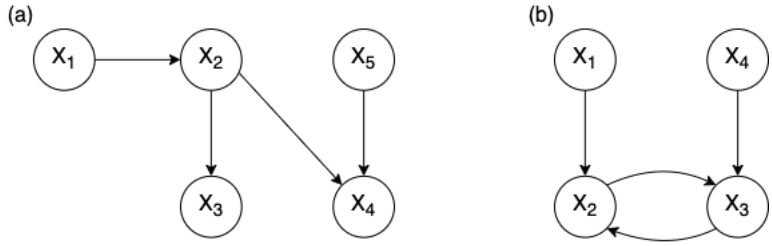
In this section, we introduce the basic concepts of graphical models, which are necessary to understand the causal discovery methods that we will study in the remainder of this paper.

2.1 Graphical Models

A graph ($\mathcal{G}$) is a diagram made up of a set of vertices ($\mathcal{V}$), and edges ($\mathcal{E}$) – the connections between the vertices – denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. A probabilistic graphical model uses a graph to express the conditional (in)dependencies between random variables, where the vertices represent random variables, and the edges encode conditional dependencies holding among the set of variables (Lauritzen, 1996). In a *causal* graphical model, on the other hand, the edges describe *causal* relationships between variables; the edges are typically directed, with $A \to B$ denoting that intervening on A results in a change in the probability distribution of B (Geiger & Pearl, 1990). Two example causal graphical models are shown in Figure 1. Figure 1a does not contain any cycles, whereas Figure 1b contains a cycle. Hence, they are called a *directed acyclic graph* (DAG) and *directed cyclic graph* (DCG), respectively.

Causal graphical models also describe patterns of statistical (in)dependencies, which can be read off the structure of the graph using Pearl's *d-separation criterion* (Geiger, Verma, & Pearl, 1990). For instance, in Figure 1a, we see a *chain* structure $X_1 \to X_2 \to X_3$, which implies that X_1 and X_3 are marginally dependent ($X_1 \not\perp\!\!\!\perp X_3$), but independent conditional on X_2 ($X_1 \perp\!\!\!\perp X_3 \mid X_2$). More formally, we would say X_1 and X_3 are *d-seperated* by X_2. A *fork* structure $X_3 \leftarrow X_2 \to X_4$ implies the same pattern of (in)dependencies; X_3 and X_4 are marginally dependent ($X_3 \not\perp\!\!\!\perp X_4$), but independent conditional on X_2 ($X_3 \perp\!\!\!\perp X_4 \mid X_2$). That is, X_3 and X_4 are *d-connected* given an empty set, but *d-separated* by X_2. However, a *collider* structure $X_2 \to X_4 \leftarrow X_5$ implies a contrasting pattern; here X_2 and X_5 are marginally independent ($X_2 \perp\!\!\!\perp X_5$), but dependent conditional on X_4 ($X_2 \not\perp\!\!\!\perp X_5 \mid X_4$). This distinguishing characteristic of colliders is crucial when identifying the directions of causal relations, as will be shown later in section 2.3.

Figure 1. Example causal graphical models.



Note. (a) is the example directed acyclic graph (DAG). (b) is the example directed cyclic graph (DCG).

2.2 Acyclic vs. Cyclic Causal Graphs

The *d-separation criterion* described above applies to all acyclic graphs, but applies to graphs with cycles only under certain conditions. To understand these conditions, first we need to introduce some graph terminology. In the field of graphical models, we use kinship terminology to describe the structure of a graph as follows:

$$\text{If } \begin{Bmatrix} A \to B \\ A \leftarrow B \\ A \to \cdots \to B \text{ or } A = B \\ A \leftarrow \cdots \leftarrow B \text{ or } A = B \end{Bmatrix} \text{ in } \mathcal{G} \text{ then } A \text{ is a } \begin{Bmatrix} \text{parent} \\ \text{child} \\ \text{ancestor} \\ \text{descendant} \end{Bmatrix} \text{ of } B \text{ and } \begin{Bmatrix} A \in pa_{\mathcal{G}}(B) \\ A \in ch_{\mathcal{G}}(B) \\ A \in an_{\mathcal{G}}(B) \\ A \in de_{\mathcal{G}}(B) \end{Bmatrix}.$$

Also, when there exists an edge between two vertices $A - B$, we say A and B are *adjacent*. For example, in Figure 1b, $X_1 \in pa_{\mathcal{G}}\, X_2$, $X_2 \in ch_{\mathcal{G}}\, X_1$, $\{X_1, X_2, X_3, X_4\} \in an_{\mathcal{G}}\, X_3$, $\{X_1, X_2, X_3\} \in de_{\mathcal{G}}\, X_1$, and X_2 is adjacent to X_1 and X_3. With this in place, we can define the *global Markov* condition, which states that d-separation relations represented in causal graphs can be used to read off statistical independence relations, such that:

$$\text{if } A \perp\!\!\!\perp_{\mathcal{G}} B \mid C \implies X_A \perp\!\!\!\perp X_B \mid X_C \text{ for all subsets of } A, B, C,$$

where $\perp\!\!\!\perp_{\mathcal{G}}$ refers to d-separation. If the causal graph is *acyclic* (i.e., DAG), then the *global Markov* condition holds regardless of the functional form of the causal relations and the distributions of the variables involved (Lauritzen, 1996). Furthermore, in DAGs, the *global Markov* property also entails the *local Markov* property, which states that a variable is independent of its non-descendants given its parents (Lauritzen, 2000). The fact that one Markov property instantly implies the other comes in handy when reading off conditional (in)dependencies from a graph.

In *cyclic* graphs (i.e., DCG), however, the situation is not so straightforward. In DCGs, the global Markov property does not always hold. Spirtes (1994) showed that it holds for DCGs when the causal relations are *linear* and the error terms are *independent*. In addition, the local Markov property may not hold, even when the global Markov property holds. For example, in Figure 1b, the global Markov property is preserved ($X_1 \perp\!\!\!\perp_{\mathcal{G}} X_4 \mid \{X_2, X_3\} \implies X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$), but the local

Markov property is violated as $X_2 \not\perp\!\!\!\perp_{\mathcal{G}} X_4 \mid X_3$ (i.e., X_2 is *not* independent of its non-descendant X_4 given its parent X_3). This is because X_3 is a parent of X_2 and a collider at the same time ($X_2 \rightarrow X_3 \leftarrow X_4$).

Accordingly, we limit the scope of our study to cyclic causal graphs that represent *linear* causal relationships with *independent* error terms, so for which the global Markov property is satisfied. Additionally, we make use of one more assumption, known as *faithfulness*, which is required for constraint-based causal discovery. *Faithfulness* assumption is essentially the reverse of the global Markov property, stating that statistical independencies map onto the structure of causal graphs:

$$X_A \perp\!\!\!\perp X_B \mid X_C \implies A \perp\!\!\!\perp_{\mathcal{G}} B \mid C.$$

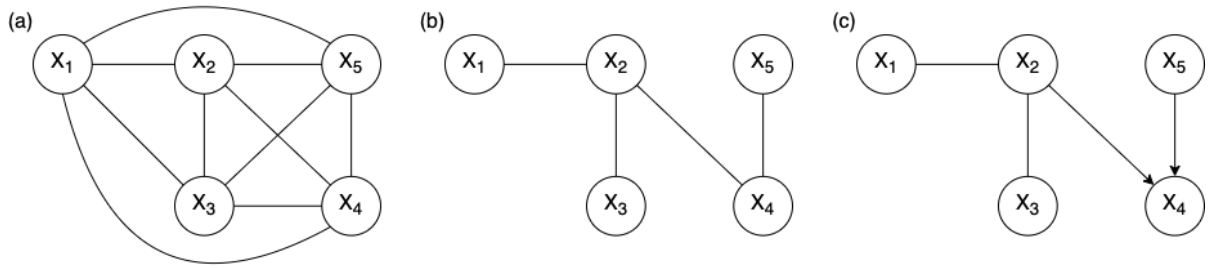
Together with the global Markov property, faithfulness enables us to make inferences about patterns of directed causal relationships in graphs by testing for statistical independencies among variables (Bongers, Forré, Peters, & Mooij, 2021).

2.3 Causal Discovery Primer

Under the aforementioned assumptions, constraint-based causal discovery methods seek to retrieve a graph structure that corresponds exactly to the conditional independencies found in observational data. By constraint-based, we mean a two-step procedure in which an estimated set of conditional independencies among variables is (1) first used to establish the *skeleton* – an undirected version of the underlying causal graph – and (2) second, more detailed independence patterns are used to give directions to the edges. In general, constraint-based techniques are unable to uniquely identify the underlying causal graph, but instead identify a set of causal graphs that imply the same statistical independence relations.

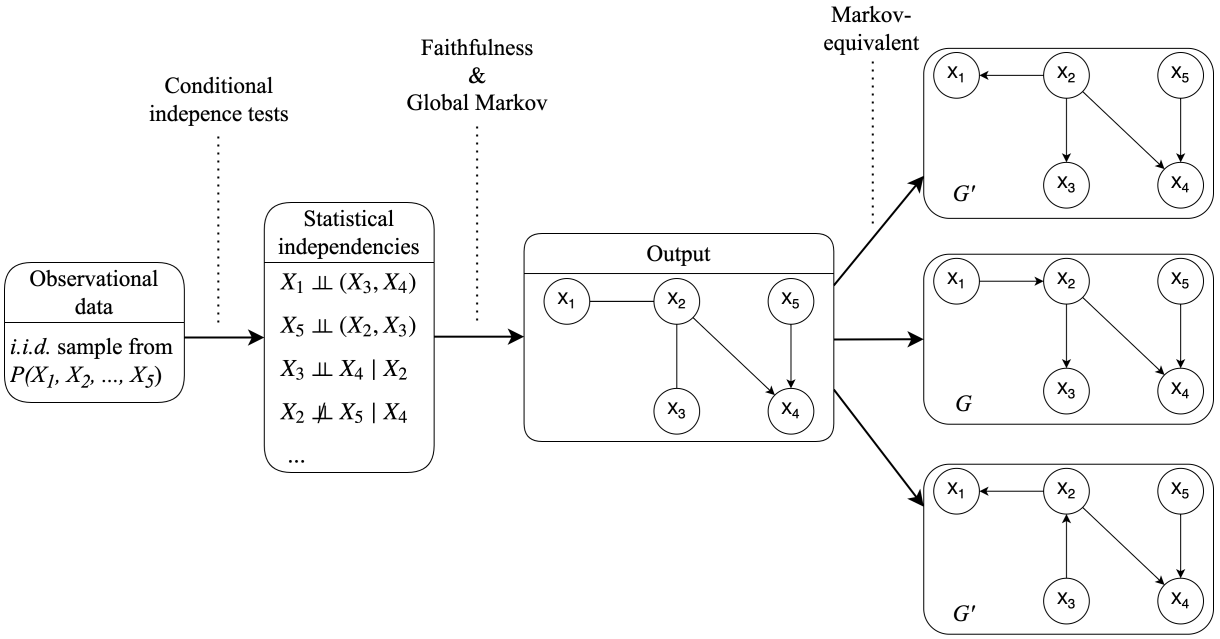
To develop an intuition for this, we examine how a constraint-based method works for a relatively simple case of the DAG from Figure 1a. First, we start with a fully-connected graph as shown in Figure 2a. Then, the *skeleton* is estimated by searching for conditional independencies; if two variables are independent when conditioning on any subset of the remaining variables (e.g., $X_5 \perp\!\!\!\perp (X_1, X_2, X_3)$, $X_1 \perp\!\!\!\perp X_3 \mid X_2$, $X_1 \perp\!\!\!\perp X_4 \mid X_2$, ...), then the edge between the two variables is removed (see Figure 2b). In the second step, the algorithm searches for *colliders* that induce the distinctive patterns of (in)dependencies (e.g., $X_2 \perp\!\!\!\perp X_5$ and $X_2 \not\perp\!\!\!\perp X_5 \mid X_4$) to determine directions of edges (e.g., $X_2 \rightarrow X_4 \leftarrow X_5$; see Figure 2c). Note that the resulting graph in Figure 2c is not identical to the original true graph $\mathcal{G}$, as the two edges between $X_1 - X_2$ and $X_2 - X_3$ remain undirected. There are in fact three DAGs that are implied by the resulting graph, as shown in Figure 3. These DAGs are called *Markov equivalent*, meaning that they encode the same set of conditional independencies (i.e., same d-separation relations hold). This illustrates a general difficulty in constraint-based methods; there are usually multiple causal graphs consistent with the found conditional independencies, which form a *Markov equivalence class*, $Equiv(\mathcal{G})$. See Figure 3 for the summary of the constraint-based approach procedure.

Figure 2. Steps of a constraint-based algorithm.



Note. (a) shows the fully-connected graph, which is the starting point. (b) shows the estimated *skeleton* – an undirected graph of the underlying causal structure – after the first step. (c) shows the resulting graph after the second step, which represents a *Markov equivalence class* of DAGs (i.e., a set of DAGs that encode the same set of conditional independencies).

Figure 3. Summary of the constraint-based causal discovery procedure.



Note. First, a constraint-based algorithm performs a series of conditional independence tests on observational (*i.i.d.*: independent and identically distributed) data . Under the faithfulness and global Markov assumption, the algorithm estimates a graph structure based on the statistical independence patterns. The output is a partially directed graph (as some edges remain undirected). It can represent multiple graphs that are *Markov equivalent*, meaning that they imply the same statistical independencies. This equivalent set of graphs is called *Markov equivalence class* and in this example, it consists of three different DAGs including the true DAG (*G*).

3 Methods

In this paper, we compare the performance of three different constraint-based algorithms for cyclic models using a simulation study: *cyclic causal discovery* (CCD) (Richardson, 1996b), *fast causal inference* (FCI) (Mooij & Claassen, 2020), and *cyclic causal inference* (CCI) (Strobl, 2019). In this section, we provide a detailed description on one of the considered algorithms, CCD, and illustrate how we conduct the simulation study. The other two algorithms, FCI and CCI, work almost the same way, except they do not require the assumption of no unobserved confounding needed by the CCD algorithm. See Appendices for the specifics of each of the three algorithms.

3.1 CCD Algorithm

In what follows, we explain how the CCD algorithm works by describing the corresponding output and step-by-step tracing the algorithm.

3.1.1 Output Representation: Partial Ancestral Graph (PAG)

Previously in section 2.3, we showed how a constraint-based method estimated a Markov equivalence class of DAGs. Here, we are interested in directed *cyclic* graphs (DCG). As was the case with DAGs, there typically exist multiple DCGs that imply the same statistical independencies. To represent a Markov equivalence class of DCGs, we use a *partial ancestral graph* (PAG). Now that we allow cycles, the causal semantics of edges in PAGs are more complicated; directed edges denote causal *ancestry* (i.e., $A \rightarrow B$ means A is an *ancestor* of B), and there are three different types of edge-endpoints available: $\{\circ, >, -\}$. Additionally, either a *solid underlining* or *dotted underlining* can be added in a PAG. In the following definition that provides the semantics for PAGs (Richardson, 1996b), $*$ is used as a *meta-symbol* implying one of the three possible edge-endpoints.

Definition 1 (PAG) Ψ is a PAG for the directed cyclic graph $\mathcal{G}$ iff:

1. There is an edge between A and B in Ψ iff A and B are d-connected in $\mathcal{G}$ given all subsets of the remaining vertices.
2. If there is an edge $A - *B$ in Ψ, it means A is an ancestor of B in every graph in the equivalent set, $Equiv(\mathcal{G})$.
3. If there is an edge $A * -> B$ in Ψ, it means B is *not* an ancestor of A in $Equiv(\mathcal{G})$.
4. If there is a solid underlining $A * - * \underline{B} * - * C$ in Ψ, then B is an ancestor of (at least one of) A or C in $Equiv(\mathcal{G})$.
5. When $A \rightarrow B \leftarrow C$, a dotted underlining is added $A \rightarrow \underline{\underline{B}} \leftarrow C$ iff B is *not* a descendant of a common child of A and C in $Equiv(\mathcal{G})$.
6. Any edge-endpoints not marked in one of the ways above are left with a circle: $\circ - *$, indicating that the algorithm is unsure of the direction.

3.1.2 Steps of CCD Algorithm

The CCD algorithm consists of 6 steps. We illustrate each step of CCD using the example DCG from Figure 4a.[1] The algorithm initially starts with a fully-connected PAG with circle endpoints ($\circ\!\!-\!\!\circ$), as shown in Figure 4b.

Step 1. Estimate the skeleton based on conditional independencies. When two vertices A and B are *d-separated* given a set S, remove $A * - * B$ and record $S = \mathbf{Sepset}\langle A, B\rangle = \mathbf{Sepset}\langle B, A\rangle$. Since $X_1 \perp\!\!\!\perp X_4 \mid \emptyset$ in our example DCG, $X_1 \circ\!\!-\!\!\circ X_4$ is removed and $\mathbf{Sepset}\langle X_1, X_4\rangle = \mathbf{Sepset}\langle X_4, X_1\rangle = \emptyset$ is recorded, resulting in Figure 4c.

Step 2. Search for collider structures. If $B \notin \mathbf{Sepset}\langle A, C\rangle$ in a triplet $A * - * B * - * C$, identify B as a collider and orient $A \rightarrow B \leftarrow C$. Given that $X_2 \notin \mathbf{Sepset}\langle X_1, X_4\rangle$ and $X_3 \notin \mathbf{Sepset}\langle X_1, X_4\rangle$ in our example, $X_1 \circ\!\!-\!\!\circ X_2 \circ\!\!-\!\!\circ X_4$ and $X_1 \circ\!\!-\!\!\circ X_3 \circ\!\!-\!\!\circ X_4$ are oriented respectively as $X_1 \rightarrow X_2 \leftarrow X_4$ and $X_1 \rightarrow X_3 \leftarrow X_4$, resulting in Figure 4d.

Step 3. Check for additional d-separating relations in each triplet $\langle A, B, C\rangle$ such that:
(i) A is not adjacent to B or C, *(ii)* B and C are adjacent, and *(iii)* $B \notin \mathbf{Sepset}\langle A, C\rangle$. If such triplets exist, orient $B * - * C$ as $B \leftarrow C$. No such triplets are found in our example, hence no further orientations are performed in step 3.

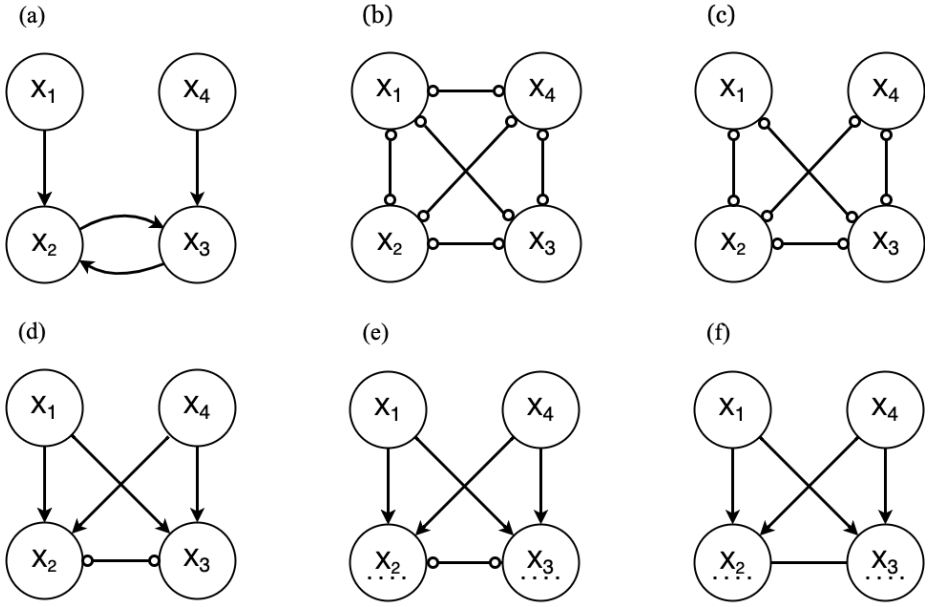
Step 4. Search for **Supsets**, which are d-separating sets including colliders. For each collider structure $A \rightarrow B \leftarrow C$, check if there is any set T including B that d-separates A and C. When such exists, record $T = \mathbf{Supset}\langle A, B, C\rangle$ and add a dotted-underlining $A \rightarrow \underline{B} \leftarrow C$. Since $X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$ in our example, $\mathbf{Supset}\langle X_1, X_2, X_4\rangle = \mathbf{Supset}\langle X_1, X_3, X_4\rangle = \{X_2, X_3\}$ is recorded and each of the colliders is dotted-underlined as $X_1 \rightarrow \underline{X_2} \leftarrow X_4$ and $X_1 \rightarrow \underline{X_3} \leftarrow X_4$, resulting in Figure 4e.

Step 5. Search for quadruplets – four ordered vertices $\langle A, B, C, D\rangle$ – where:
(i) $A \rightarrow \underline{B} \leftarrow C$, *(ii)* $A \rightarrow D \leftarrow C$ or $A \rightarrow \underline{D} \leftarrow C$, and *(iii)* B and D are adjacent. If $D \in \mathbf{Supset}\langle A, B, C\rangle$ in such quadruplets, orient $B * - * D$ as $B * \!\!-\!\! D$. Else, orient $B * - * D$ as $B \rightarrow D$. In our example, there is a quadruplet such that *(i)* $X_1 \rightarrow \underline{X_2} \leftarrow X_4$, *(ii)* $X_1 \rightarrow \underline{X_3} \leftarrow X_4$, and *(iii)* X_2 and X_3 are adjacent. Since $X_2 \in \mathbf{Supset}\langle X_1, X_3, X_4\rangle$ and $X_3 \in \mathbf{Supset}\langle X_1, X_2, X_4\rangle$, $X_2 \circ\!\!-\!\!\circ X_3$ is oriented as $X_2 \multimap X_3$, then $X_2 \multimap X_3$ is subsequently oriented as $X_2 - X_3$, which results in Figure 4f.

Step 6. Search for quadruplets $\langle A, B, C, D\rangle$, where $A \rightarrow \underline{B} \leftarrow C$ while D is adjacent to neither A nor C. If A and D are d-connected given $\mathbf{Supset}\langle A, B, C\rangle \cup D$, then orient $B* \multimap D$ as $B \rightarrow D$. In our example, there exist no such quadruplets, therefore Figure 4f remains as the final PAG. As shown in Figure 5, the resulting PAG implies two different DCGs including the true graph $\mathcal{G}$, which form a Markov equivalence class.

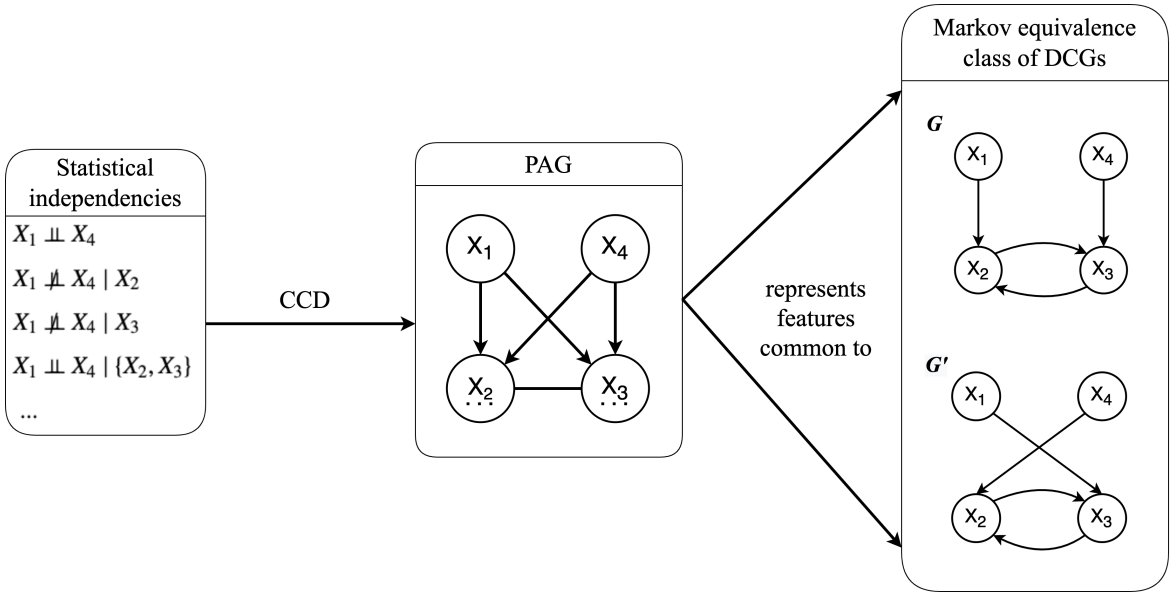
[1] It is identical to the example DCG that we previously introduced in Figure 1b.

Figure 4. Trace of CCD algorithm.



Note. (a) shows the true directed cyclic graph. (b) shows the fully-connected graph, which is the initial state of the PAG. (c) shows the estimated skeleton (i.e., an undirected graph of the underlying structure) after step 1. (d) shows the state of the PAG after step 2, where some of the edges are oriented given the identified colliders. (e) shows the state of PAG after step 4, where the *Supsets* are identified and the corresponding colliders are dotted-underlined. (f) shows the final PAG after step 5, where an additional edge between X_2 and X_3 is oriented.

Figure 5. Summary of CCD algorithm.



Note. Given the found statistical independencies, CCD constructs a partial ancestral graph (PAG), which represents the features that are common to every directed cyclic graph (DCG) in a Markov equivalence class. In this example, the Markov equivalence class consists of two different DCGs, including the true graph $\mathcal{G}$.

$$X_1 \perp\!\!\!\perp X_4 \tag{1}$$

$$X_1 \not\perp\!\!\!\perp X_4 \mid X_2 \tag{2}$$

$$X_1 \not\perp\!\!\!\perp X_4 \mid X_3 \tag{3}$$

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\} \tag{4}$$

$$\ldots \tag{5}$$

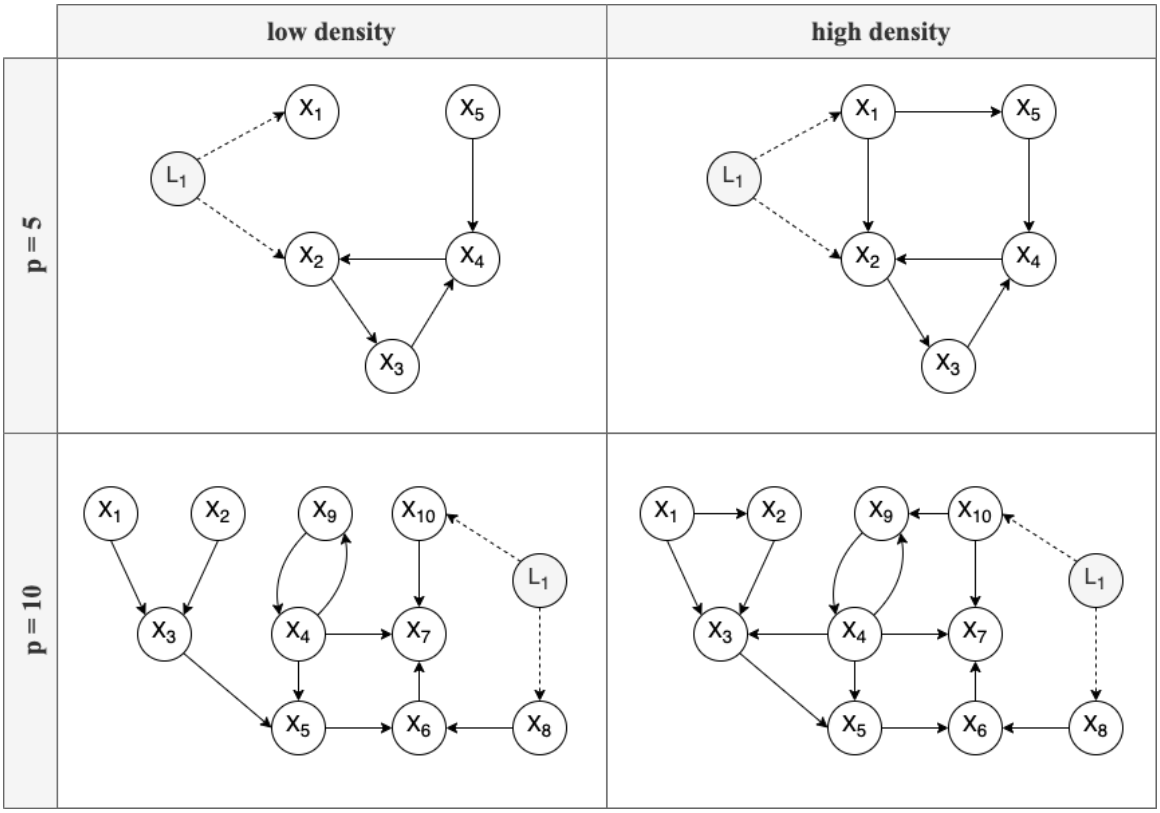
3.2 Simulation

To evaluate the performance of the considered algorithms, we perform a simulation study. In this section, we discuss the simulation design, data generating process, and evaluation metrics in detail.

3.2.1 Simulation Design

We test each algorithm under different conditions by varying the number of variables (rows of Figure 6) and the number of edges – the density (columns of Figure 6). We also evaluate the effect of an unobserved confounder by adding a latent variable (L_1 in Figure 6). Lastly, we vary the sample size across the range we often encounter in psychological research; $n \in \{150, 500, 1000\}$ for every simulated cyclic model (Constantin & Cramer, 2022). Thus, it leads to a $2 \times 2 \times 2 \times 3$ design; number of variables $\times$ density $\times$ latent confounder (presence/absence) $\times$ sample size.

Figure 6. Simulation settings.



Note. We vary the number of variables: $p \in \{5, 10\}$, density: high / low, a latent confounder (L_1): absence / presnece, and the sample size: $n \in \{150, 500, 1000\}$, which leads to a $2 \times 2 \times 2 \times 3$ simulation design.

3.2.2 Generating Data

As illustrated above, we simulate data from different cyclic causal models, all of which are characterized by *linear* relations and *independent Gaussian* error terms. Not only that they are commonly assumed in psychological research, but also under these assumptions, the requirement for using causal discovery methods – global Markov property – holds, as shown in section 2.2.

To generate data, we first define a coefficient matrix $\mathbf{B}$ and sample the error terms (ϵ) from the specified joint Gaussian distribution. After drawing the values of ϵ, we can obtain the data by solving the following systems of equations: $\mathbf{X} = (\mathbf{I} - \mathbf{B})^{-1}\varepsilon$, where $\mathbf{I}$ denotes the identity matrix and ε denotes a set of jointly independent Gaussian error terms. Note that the equations may not have unique solutions for some of the cyclic causal models. The necessary condition to have a unique solution is that $(\mathbf{I} - \mathbf{B})$ is invertible. In cyclic models, this condition is only satisfied when the absolute values of eigenvalues of $\mathbf{B}$ are smaller than one, $|\lambda| < 1$ (Eberhardt, Hoyer, & Scheines, 2010). Hence, we ensure that the specified $\mathbf{B}$ matrix satisfies the aforementioned condition.

3.2.3 Evaluation Metrics

We assess the performance using both *local* and *global* evaluation metrics; at a local level, we look at the individual edge-endpoints and at a global level, we look at the graph structure as a whole. As the local metrics, we utilize *precision*, *recall*, and *uncertainty rate*. Precision reflects the prediction accuracy (i.e., out of all predicted cases, how many are correct), and recall reflects the retrieval rate (i.e., out of all true cases, how many are retrieved). There are in total four possibilities for each edge-endpoint in a resulting graph: no edge-endpoint (null), arrow head (>), arrow tail (−), and circle (∘). Given that a circle endpoint implies that the algorithm is unsure of the direction, the uncertainty rate is defined as the proportion of the circle endpoints occurred in the output. For the other endpoints, we calculate the precision and recall. For example, for the arrow head, they can be computed as: $precision = \frac{a}{a+d+g}$ and $recall = \frac{a}{a+b+c}$ (see Table 1).

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{ False Positive}}$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive } + \text{ False Negative}}$$

$$Uncertainty = \frac{\text{Number of circle endpoint } (\circ)}{\text{Total number of edge-endpoints}}$$

As the global metric, we use *structural Hamming distance* (SHD) (de Jongh & Druzdzel, 2009). SHD quantifies the level of differences between two graphs by counting the number of edge insertions, deletions, and direction changes required to move from one graph (estimated graph $\hat{\mathcal{G}}$) to the other (true graph $\mathcal{G}$). It can be formulated as: $SHD = A + D + C$, where A represents the number of added edges, D represents the number of deleted edges, and C represents the number of direction changes. Thus, the smaller the SHD value is, the more similar $\hat{\mathcal{G}}$ is to $\mathcal{G}$, indicating that the algorithm recovered the true graph well.

Table 1. Confusion matrix for three types of edge-endpoints.

		Estimated endpoint		
		Arrow head (>)	Arrow tail (−)	Null
True endpoint	Arrow head (>)	*a*	*b*	*c*
	Arrow tail (−)	*d*	*e*	*f*
	Null	*g*	*h*	*i*

Note. The true endpoints are presented in rows, and the estimated endpoints are presented in columns. In total, there are four possible edge-endpoints that can occur: arrow head, arrow tail, null (no endpoint), and circle. The circle endpoint (∘), however, is not counted towards true/false positives/negatives, but used to compute the *uncertainty rate*.

References

Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_15

Bongers, S., Forré, P., Peters, J., & Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, *49*(5), 2885 – 2915. https://doi.org/10.1214/21-AOS2064

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91-121. (PMID: 23537483) https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., ... Waldorp, L. J. (2021, August). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 1–18. https://doi.org/10.1038/s43586-021-00055-w

Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on bayesian networks for psychopathology researchers. *Psychological Methods*. https://doi.org/10.1037/met0000479

Constantin, M., & Cramer, A. O. J. (2022). Sample size recommendations for estimating cross-sectional network models. *OSF*. https://doi.org/10.17605/OSF.IO/ZKAXU

Dablander, F., & Hinne, M. (2019, May). Node centrality measures are a poor substitute for causal inference. *Scientific Reports*, *9*(1), 6846. https://doi.org/10.1038/s41598-019-43033-9

de Jongh, M., & Druzdzel, M. J. (2009). A comparison of structural distance measures for causal bayesian network models. *Recent Advances in Intelligent Information Systems, Challenging Problems of Science, Computer Science series*, 443–456.

Eberhardt, F., Hoyer, P., & Scheines, R. (2010, 13–15 May). Combining experiments to discover linear cyclic models with latent variables. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (Vol. 9, pp. 185–192). PMLR.

Forré, P., & Mooij, J. M. (2017). Markov Properties for Graphical Models with Cycles and Latent Variables. *arXiv preprint arXiv:1710.08775*. https://doi.org/10.48550/arXiv.1710.08775

Geiger, D., & Pearl, J. (1990). On the logic of causal models. In R. D. Shachter, T. S. Levitt, L. N. Kanal, & J. F. Lemmer (Eds.), *Machine intelligence and pattern recognition* (Vol. 9, pp. 3–14). North-Holland. https://doi.org/10.1016/B978-0-444-88650-7.50006-8

Geiger, D., Verma, T., & Pearl, J. (1990). d-Separation: From Theorems to Algorithms. In M. Henrion, R. D. Shachter, L. N. Kanal, & J. F. Lemmer (Eds.), *Machine Intelligence and Pattern Recognition* (Vol. 10, pp. 139–148). North-Holland. https://doi.org/10.1016/B978-0-444-88738-2.50018-X

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

Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2021, November). Modeling psychopathology: From data models to formal theories. *Psychological Methods*. https://doi.org/10.1037/met0000303

Kossakowski, J., Waldorp, L. J., & van der Maas, H. L. J. (2021). The search for causality: A comparison of different techniques for causal inference graphs. *Psychological Methods*, *26*(6), 719–742. https://doi.org/10.1037/met0000390

Lauritzen, S. (1996). *Graphical models* (Vol. 17). Clarendon Press.

Lauritzen, S. (2000). *Graphical models for causal inference.* Complex Stochastic Systems. London/Boca Raton: Chapman and Hall/CRC Press.

Mooij, J. M., & Claassen, T. (2020, Aug). Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In J. Peters & D. Sontag (Eds.), *Proceedings of the 36th conference on uncertainty in artificial intelligence (uai)* (Vol. 124, pp. 1159–1168). PMLR.

Pearl, J. (2010). Causal Inference. In *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008* (pp. 39–58). PMLR.

Richardson, T. (1996a). *Discovering cyclic causal structure*. Carnegie Mellon [Department of Philosophy].

Richardson, T. (1996b). A discovery algorithm for directed cyclic graphs. In *Proceedings of the twelfth international conference on uncertainty in artificial intelligence* (p. 454–461). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal*, *0*(0), 1-18. https://doi.org/10.1080/10705511.2022.2056039

Spirtes, P. (1993). Directed cyclic graphs, conditional independence, and non-recursive linear structural equation models..

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

Spirtes, P., & Glymour, C. (1991). An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review*, *9*(1), 62–72. (Publisher: SAGE Publications Inc) https://doi.org/10.1177/089443939100900106

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.

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. https://doi.org/10.1007/s41060-018-0158-2

Wittenborn, A. K., Rahmandad, H., Rick, J., & Hosseinichimeh, N. (2016). Depression as a systemic syndrome: mapping the feedback loops of major depressive disorder. *Psychological Medicine*, *46*(3), 551–562. https://doi.org/10.1017/S0033291715002044