Discovering Cyclic Causal Models in Psychological Research

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

[1]Department of Methodology and Statistics, Utrecht University

January 16, 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: 2499

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 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 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 confounders (Dablander & Hinne, 2019; Ryan, Bringmann, & Schuurman, 2022).

In the field of causal discovery, using statistical independencies estimated from observational data to infer causal structures is known as *constraint-based* causal discovery (Spirtes & Glymour, 1991). Ryan et al. (2022) suggest that network models could be replaced by purpose-built constraint-based causal discovery methods. However, the most popular and well-studied constraint-based methods 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). For example,

*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

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* methods. Although some cyclic causal discovery algorithms have been developed (Mooij & Claassen, 2020), they have not been as well studied as their acyclic counterparts, which is in part 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 cyclic causal discovery methods in psychology, and much remains unknown about the performance of these methods.

Therefore, in this paper, we aim to address the following question: how well do cyclic causal discovery methods perform in typical psychological research contexts? The goal of this study is twofold. First, we provide an accessible overview of different cyclic causal discovery methods. Second, we investigate how well each of these methods works by means of a simulation study.

2 Background

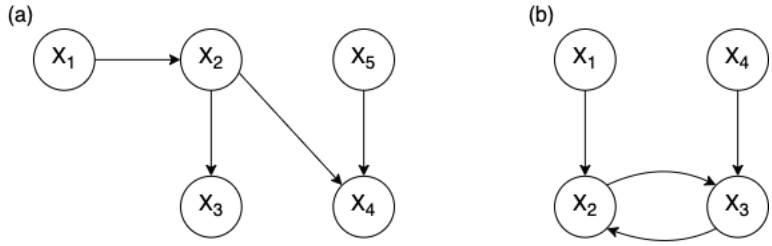
In this section, we introduce the basic concepts of graphical models, which are necessary to understand the causal discovery methods that we will study in the remainder of this paper.

2.1 Graphical Models

A graph ($\mathcal{G}$) is a diagram made up of a set of vertices ($\mathcal{V}$) and edges ($\mathcal{E}$), describing connections between 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 does, hence called a *directed acyclic graph* (DAG) and *directed cyclic graph* (DCG), respectively.

Causal graphical models also describe patterns of statistical independencies, which can be read off from 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-separated* by X_2. A *fork* structure $X_3 \leftarrow X_2 \to X_4$ implies the same pattern of independencies; 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 only applies to graphs with cycles under certain conditions. To understand these conditions, first we need to introduce some graph terminology. We can use kinship terminology to describe a graph structure 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 $A - B$, A and B are said to be *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 causal graphs are *acyclic* (i.e., DAG), then the *global Markov* condition holds regardless of the functional forms of causal relations and the distributions of variables involved (Lauritzen, 1996). In addition, in DAGs, the *global Markov* condition also entails the *local Markov* condition stating 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 independencies from a graph.

In contrast to the acyclic case, the situation is not so straightforward in *cyclic* graphs (i.e., DCG). In DCGs, the global Markov property does not always hold. Spirtes (1994), in fact, showed that it holds for DCGs when causal relations are *linear* and error terms are *independent*. Furthermore, 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 both a parent of X_2 and a collider ($X_2 \rightarrow X_3 \leftarrow X_4$) at the same time.

Accordingly, we limit the scope of our study to cyclic causal graphs that represent *linear* causal relationships with jointly *independent* error terms, so for which the global Markov condition 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 condition, stating that statistical independencies map onto the structure of causal graphs such that:

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

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

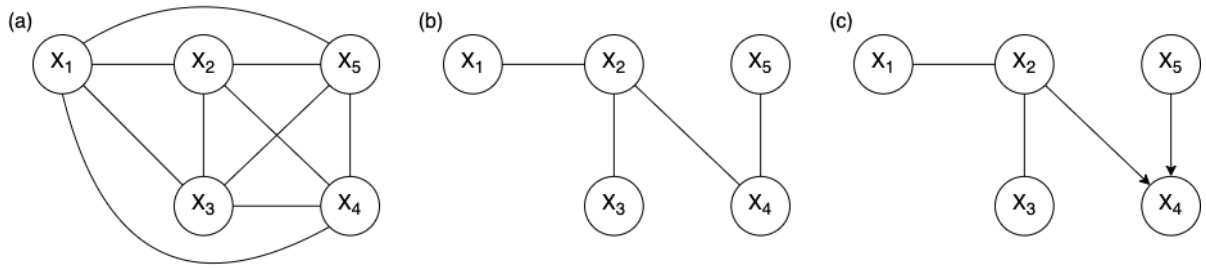
2.3 A Primer on Constraint-Based Causal Discovery

Under the aforementioned assumptions, constraint-based methods seek to recover the underlying causal structure using conditional independencies estimated from observational data. Constraint-based methods typically employ a two-step procedure; (1) first, establishing the *skeleton* – an undirected version of the underlying graph – and (2) second, attempting to assign directions to the edges. In general, constraint-based techniques are unable to uniquely identify the underlying causal graph, but instead return 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 the relatively simple DAG from Figure 1a. We start with a fully-connected graph, as shown in Figure 2a. In the first step, the *skeleton* is estimated by testing for conditional independence; if two variables are independent when conditioning on any subset of the remaining variables (e.g., $X_1 \perp\!\!\!\perp X_3 \mid X_2$, $X_1 \perp\!\!\!\perp X_4 \mid X_2, ...$), then the edge between the two variables is removed (see Figure 2b). In the second step, (some) edges are oriented by searching for *colliders* that induce distinctive patterns of independencies (e.g., X_4 is identified as a collider given $X_2 \perp\!\!\!\perp X_5$ and $X_2 \not\perp\!\!\!\perp X_5 \mid X_4$, thus $X_2 \rightarrow X_4 \leftarrow X_5$ is oriented; 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 conditional independencies (i.e., the same d-separation relations hold), and we call such a set of equivalent graphs a *Markov equivalence class*, denoted by $Equiv(\mathcal{G})$. This illustrates a general difficulty in constraint-based approach; there are usually multiple graphs that are consistent with an observed set of statistical independencies. Here, we described the constraint-based approach procedure for DAGs, but the same principles are adapted and used for cyclic graphs, which will be the focus of the remainder of this paper.[1]

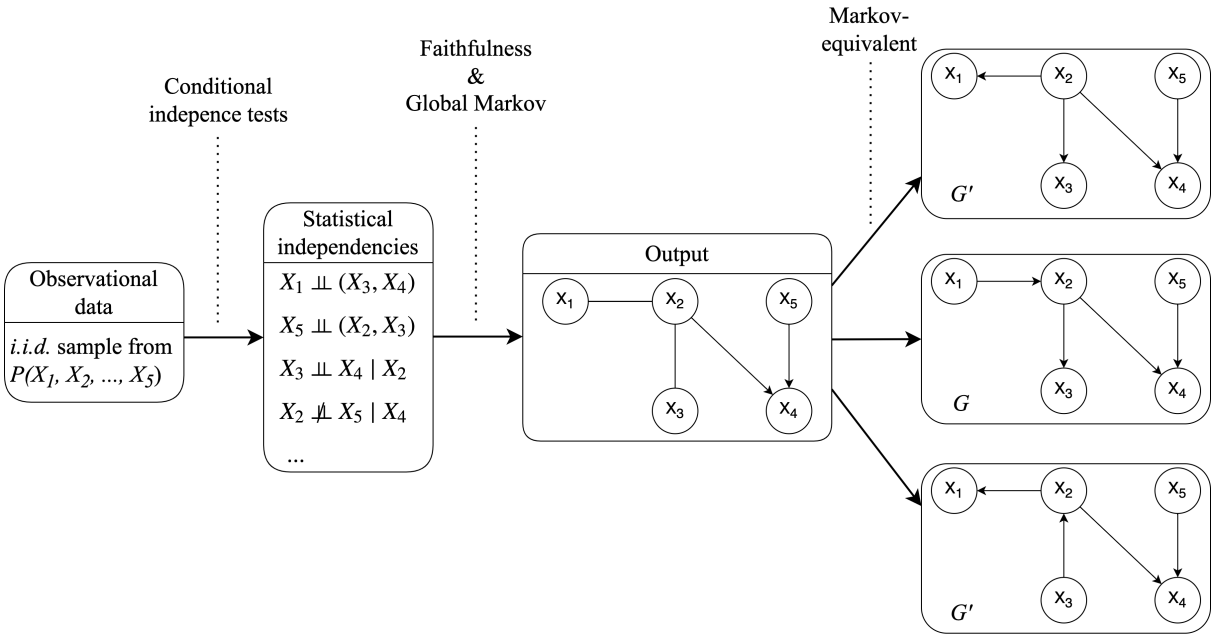
[1] See Figure 3 for a summary of the constraint-based approach procedure.

Figure 2. Steps of a constraint-based method.



Note. (a) shows the fully-connected graph for the example DAG from Figure 1a, 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 the *Markov equivalence class* of DAGs (i.e., a set of DAGs that entail the same set of conditional independencies).

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



Note. A constraint-based algorithm starts with performing 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 observed 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 independence relations. 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 graphs 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 of one of the considered algorithms, CCD, and illustrate how we conduct the simulation study. The other two algorithms, FCI and CCI, work in a similar way, except that they do not require the assumption of *no unobserved confounding* needed by the CCD algorithm.[2] (See Appendices for the specifics of each of these three algorithms).

3.1 CCD Algorithm

In what follows, we introduce the type of output generated by the CCD algorithm and trace the algorithm step-by-step using an example.

3.1.1 Output Representation: Partial Ancestral Graph (PAG)

As was the case with DAGs shown in section 2.3, there typically exist multiple directed cyclic graphs (DCG) that imply the same statistical independencies, and so are statistically indistinguishable from one another. To represent a set of equivalent DCGs, a *partial ancestral graph* (PAG) is used. Due to the possibility of cyclic relations, the causal semantics of edges in PAGs are more complicated; directed edges denote causal *ancestry* (i.e., $A \to B$ means A is an *ancestor* of B), and there are three different types of edge-endpoints available: $\{\circ, >, -\}$. Additionally, 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* indicating one of the three possible edge-endpoints.[3]

Definition 1 (PAG) Ψ is a PAG for a 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 any subset of the remaining vertices.
2. If there is an edge $A - *B$ in Ψ, then A is an ancestor of B in every graph in an equivalence class, $Equiv(\mathcal{G})$.
3. If there is an edge $A * - > B$ in Ψ, then B is *not* an ancestor of A in every graph 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 every graph in $Equiv(\mathcal{G})$.
5. If there is a collider $A \to B \leftarrow C$, a dotted underlining is added $A \to \underline{B} \leftarrow C$ iff B is *not* a descendant of a common child of A and C in every graph in $Equiv(\mathcal{G})$.
6. Any edge-endpoint not marked in one of the above ways is left with a circle $\circ - *$.

[2] The assumption that we have measured all common causes of variables involved.

[3] $A - *B$ indicates any of the following edges: $A - B$, $A - > B$, or $A - \circ B$.

3.1.2 Steps of CCD Algorithm

The CCD algorithm consists of 6 steps. We illustrate each step using the example DCG from Figure 4a.[4] The algorithm starts with a fully-connected PAG with circle endpoints, as shown in Figure 4b, and as it proceeds (some) circles will be replaced by either an arrow head or a tail.[5]

Step 1. Estimate the *ancestral* skeleton – an undirected version of a PAG – 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 $X_1 \perp\!\!\!\perp X_4 \mid \varnothing$ in our example DCG, $X_1 \circ\!\!-\!\!\circ X_4$ is removed and $\mathbf{Sepset}\langle X_1, X_4\rangle = \mathbf{Sepset}\langle X_4, X_1\rangle = \varnothing$ 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 $B \leftarrow C$. As no additional d-separating relations are found in our example, 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.

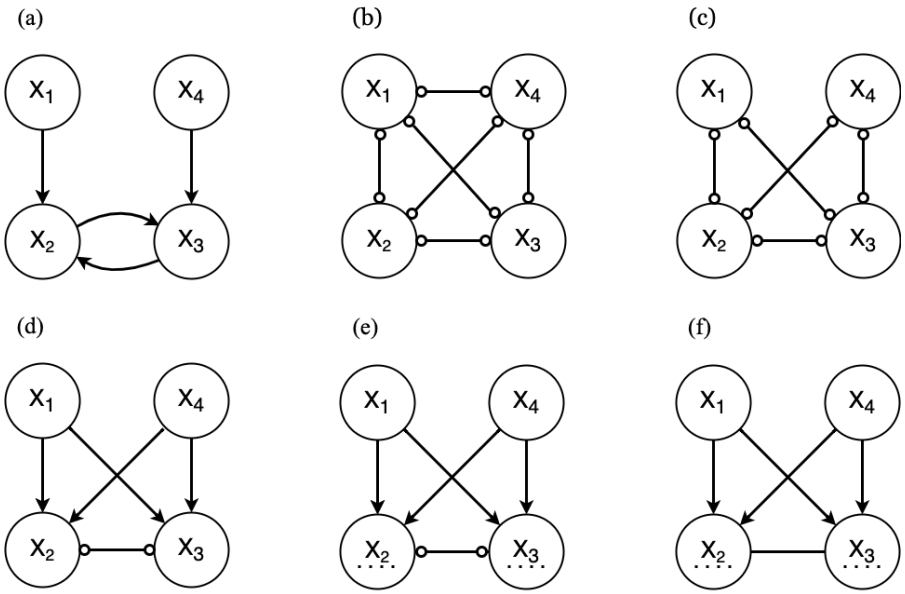
Step 5. Search for quadruplets – four ordered vertices $\langle A, B, C, D\rangle$ – where:
(i) $A \rightarrow \underline{B} \leftarrow C$, *(ii)* $A \rightarrow D \leftarrow C$ or $A \rightarrow \underline{D} \leftarrow C$, and *(iii)* B and D are adjacent. If $D \in \mathbf{Supset}\langle A, B, C\rangle$ in such quadruplets, orient $B * - * D$ as $B * \!\!- D$. Else, orient $B * - * D$ as $B \rightarrow D$. In our example, there is such a quadruplet; *(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 $X_2 \circ\!\!- X_3$, then $X_2 \circ\!\!- X_3$ is subsequently oriented as $X_2 - X_3$, resulting 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 $B* \!\!- D$ as $B \rightarrow D$. In our example, no such quadruplets exist; therefore Figure 4f remains the final PAG. As shown in Figure 5, the resulting PAG represents two different DCGs that entail the same conditional independencies.

[4] It is the same as the example DCG that we previously introduced in Figure 1b.

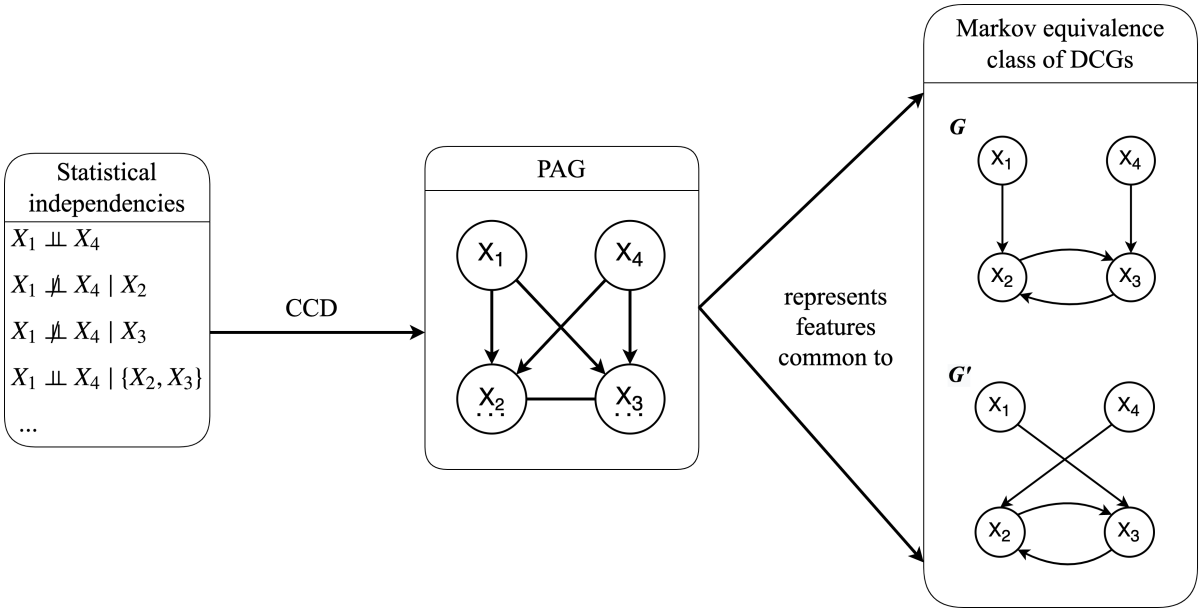
[5] It is important to note, once again, that the algorithm aims to retrieve a PAG for the underlying cyclic graph, and the edges in a PAG represent *ancestral* relations that are common to all directed cyclic graphs (DCG) in an equivalence class.

Figure 4. Trace of CCD algorithm.



Note. (a) shows the true directed cyclic graph, $\boldsymbol{G}$. (b) shows the fully-connected PAG for $\boldsymbol{G}$, which is the starting point of the algorithm. (c) shows the *ancestral* skeleton (i.e., an undirected version of the PAG) estimated in 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 the PAG after step 4, where the *Supsets* are identified and the corresponding colliders are dotted-underlined. (f) shows the final state of the PAG after step 5, where an additional edge between X_2 and X_3 is oriented.

Figure 5. Summary of CCD algorithm operation.



Note. Given the observed statistical independencies, CCD constructs a partial ancestral graph (PAG), which represents the *ancestral* 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 $\boldsymbol{G}$.

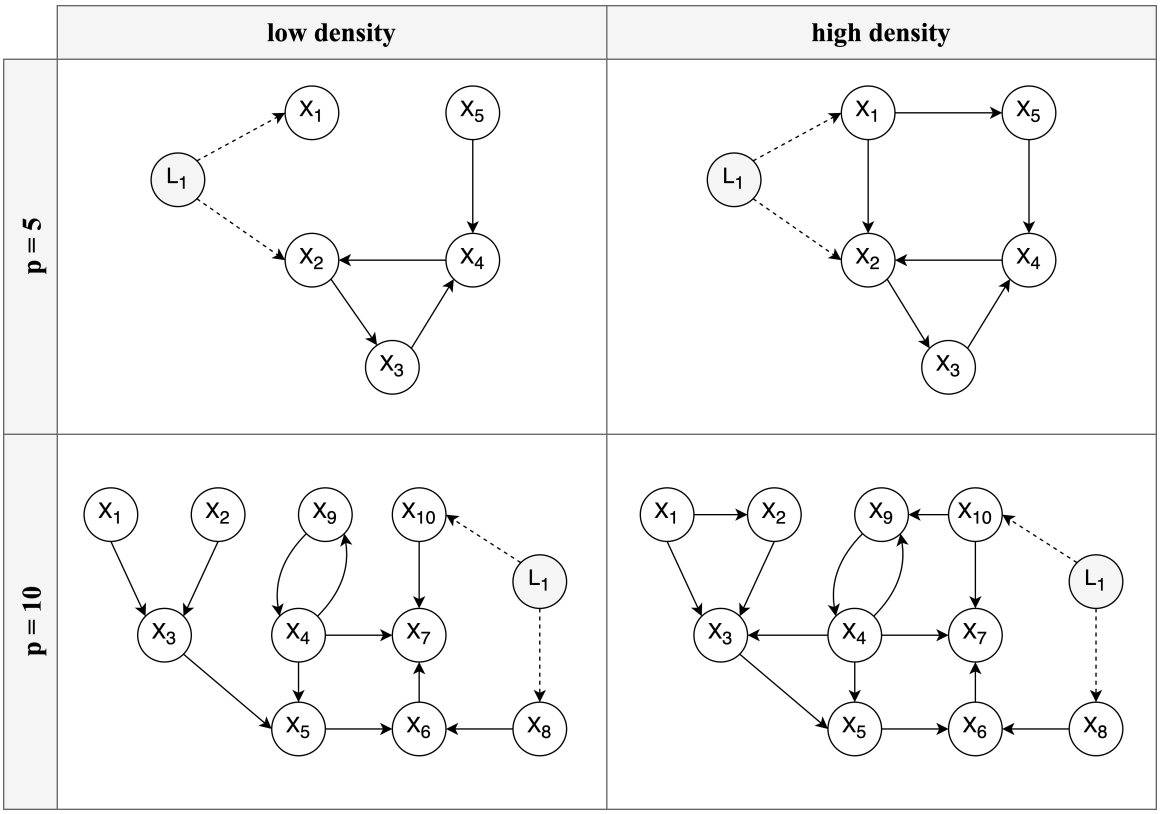
3.2 Simulation

To evaluate the performance of the considered algorithms, we conduct 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 $2\times2\times2\times3$ 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: $p \in \{5, 10\}$, the density: high / low, the influence of a latent confounder (L_1): absence / presence, and the sample size: $n \in \{150, 500, 1000\}$, which results in a $2 \times 2 \times 2 \times 3$ simulation design.

3.2.2 Data Generation

As illustrated above, we simulate data from different cyclic models, all of which are characterized by *linear* relations and *independent Gaussian* error terms. These types of models are often used in psychological research, and for such cyclic models, the global Markov property – the necessary condition for constraint-based causal discovery – also 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 independent Gaussian distributions. After drawing the values of ε, we generate observations of $\mathbf{X}$ by solving the following equation: $\mathbf{X} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\varepsilon}$, where $\mathbf{I}$ denotes the identity matrix. Note that this data generation scheme is possible provided that $(\mathbf{I} - \mathbf{B})$ is invertible, which is the case when the eigenvalues of $\mathbf{B}$ are smaller than one, $|\lambda| < 1$ (Eberhardt, Hoyer, & Scheines, 2010). While this is guaranteed if $\mathbf{B}$ defines an acyclic model, for cyclic models, this does not always hold. Therefore, we check if this condition is satisfied for every cyclic model when specifying $\mathbf{B}$ matrix.

3.2.3 Evaluation Metrics

We assess the performance of each algorithm 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*.

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

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

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

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 circle endpoints imply an algorithm is unsure of the direction of causal relations, the uncertainty rate is defined as a proportion of the circle endpoints occurred in an output. For the other endpoints, we calculate the precision and recall. For example, for the arrow head, they are computed as: $precision = \frac{a}{a+d+g}$ and $recall = \frac{a}{a+b+c}$ (see Table 1).

Table 1. Confusion matrix for the 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. There are in total four possible edge-endpoints that can occur in an output: arrow head, arrow tail, null (no endpoint), and circle. The circle endpoints (∘), however, are not counted toward the calculation of *precision* and *recall* but are used for calculating the *uncertainty rate*.

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, D, and C represent, respectively, the number of added edges, deleted edges, and direction changes. Thus, the smaller the SHD value is, the more similar $\hat{\mathcal{G}}$ is to $\mathcal{G}$, indicating that an algorithm recovers the true graph well.

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

Appendix A CCD Details

Algorithm 1 Cyclic Causal Discovery (CCD)

Input: A conditional independent oracle for a distribution $\mathcal{P}$, satisfying global directed Markov property and faithfulness conditions with respect to a directed graph $\mathcal{G}$ with vertex set $\mathcal{V}$.

Output: A PAG Ψ for the Markov equivalence class Equiv($\mathcal{G}$).

Step 1. Form a complete graph (Ψ) with the edge $\circ\!\!-\!\!\circ$ between every pair of vertices in $\mathcal{V}$.
$n = 0$
repeat
  repeat
    Select an ordered pair of variables X and Y that are adjacent in Ψ such that the number of vertices in **Adjacent**$(\Psi, X)\backslash\{Y\} \geq n$, and select a subset $\mathcal{S}$ of **Adjacent**$(\Psi, X)\backslash\{Y\}$ with n vertices.
    If $X \perp\!\!\!\perp Y \mid \mathcal{S}$, then delete the edge $X \circ\!\!-\!\!\circ Y$ and record $\mathcal{S}$ in **Sepset**$\langle X, Y\rangle$ and **Sepset**$\langle X, Y\rangle$.
  until all pairs of adjacent variables X and Y such that the number of vertices in **Adjacent**$(\Psi, X)\backslash\{Y\} \geq n$ and all sets $\mathcal{S}$ such that the number of vertices in $\mathcal{S} = n$ have been tested.
  $n = n + 1$;
until for all ordered pairs of adjacent vertices X and Y, **Adjacent**$(\Psi, X)\backslash\{Y\} < n$.
Step 2. For each triple of vertices A, B, C such that each of the pair of A, B and the pair B, C are adjacent in Ψ but the pair A, C are not adjacent in Ψ, then:
(i) orient $A * - * B * - * C$ as $A \rightarrow B \leftarrow C$ *iff* $B \notin$ **Sepset**$\langle A, B\rangle$.
(ii) orient $A * - * B * - * C$ as $A * - * \underline{B} * - * C$ *iff* $B \in$ **Sepset**$\langle A, B\rangle$.
Step 3. For each triple of vertices A, X, Y in Ψ such that (a) A is not adjacent to X or Y, (b) X and Y are adjacent, (c) $X \notin$ **Sepset**$\langle A, Y\rangle$, then orient $X * - * Y$ as $X \leftarrow Y$ if $A \not\perp\!\!\!\perp X \mid$ **Sepset**$\langle A, Y\rangle$.
Step 4. For each vertex V in Ψ form the following set: $X \in$ **Local**(Ψ, V) or there is a vertex Y such that $X \rightarrow Y \leftarrow V$ in Ψ.
$m = 0$
repeat
  repeat
    Select an ordered triple $\langle A, B, C\rangle$ such that $A \rightarrow B \leftarrow C$, A and C are not adjacent, and **Local**$(\Psi, A)\backslash\{B, C\}$ has $\geq m$ vertices.
    Select a set $T \subseteq$ **Local**$(\Psi, A)\backslash\{B, C\}$ with m vertices. If $A \perp\!\!\!\perp C \mid T \cup \{B\}$, then orient $A \rightarrow B \leftarrow C$ as $A \rightarrow \underline{B} \leftarrow C$ and record $T \cup \{B\}$ in **Supset**$\langle A, B, V\rangle$.

until for all triples such that $A \to B \leftarrow C$ (not $A \to \underline{\mathrm{B}} \leftarrow C$), A and C are not adjacent, $\mathbf{Local}(\Psi, A)\backslash\{B\}$ has $\geq m$ vertices, every subset T with m vertices has been considered.

$m = m + 1$;

until all ordered triples $\langle A, B, C\rangle$ such that $A \to B \leftarrow C$, A and C are not adjacent, are such that $\mathbf{Local}(\Psi, A)\backslash\{B\}$ have $< m$ vertices.

Step 5. If there is a quadruple A, B, C, D in Ψ of distinct vertices such that:

(i) $A \to \underline{\mathrm{B}} \leftarrow C$,

(ii) $A \to D \leftarrow C$ or $A \to \underline{\mathrm{D}} \leftarrow C$,

(iii) B and D are adjacent,

then orient $B * - * D$ ad $B \to D$ in Ψ if $D \notin \mathbf{Subset}\langle A, B, C\rangle$. Else orient $B * - * D$ as $B * - D$ in Ψ.

Step 6. For each quadruple A, B, C, D in Ψ of distinct vertices such that:

(i) D is not adjacent to both A and C,

(ii) $A \to \underline{\mathrm{B}} \leftarrow C$,

if $A \not\perp\!\!\!\perp D \mid \mathbf{Supset}\langle A, B, C\rangle \cup \{D\}$, then orient $B * - * D$ as $B \to D$ in Ψ.

Appendix B FCI Details

Algorithm 2 Fast Causal Inference (FCI)

Input: A conditional independent oracle for a distribution $\mathcal{P}$, satisfying global directed Markov property and faithfulness conditions with respect to a directed graph $\mathcal{G}$ with vertex set $\mathcal{V}$.

Output: A PAG $\hat{\mathcal{G}}'$ for the Markov equivalence class of MAGs (Equiv($\mathcal{G}$)).

Step 1. Form the complete undirected graph Q on the vertex set $\mathcal{V}$.

$n = 0$

repeat

repeat

Select an ordered pair of variables X and Y that are adjacent in Q such that the number of vertices in $\mathbf{Adjacent}(Q, X)\backslash\{Y\} \geq n$, and select a subset $\mathcal{S}$ of $\mathbf{Adjacent}(Q, X)\backslash\{Y\}$ with n vertices.

If $X \perp\!\!\!\perp Y \mid \mathcal{S}$, then delete the edge $X \circ\!\!-\!\!\circ Y$ and record $\mathcal{S}$ in $\mathbf{Sepset}\langle X, Y\rangle$ and $\mathbf{Sepset}\langle X, Y\rangle$.

until all pairs of adjacent variables X and Y such that the number of vertices in $\mathbf{Adjacent}(Q, X)\backslash\{Y\} \geq n$ and all sets $\mathcal{S}$ such that the number of vertices in $\mathcal{S} = n$ have been tested.

$n = n + 1$;

until for all ordered pairs of adjacent vertices X and Y, $\mathbf{Adjacent}(Q, X)\backslash\{Y\} < n$.

Step 2. Let Q' be the undirected graph resulting from step 1. For each triplet $\langle A, B, C\rangle$ such that each of the pair of A, B and the pair B, C are adjacent in Q' but the pair A, C are not adjacent in Q', then:

(i) orient $A * - * B * - * C$ as $A \rightarrow B \leftarrow C$ *iff* $B \notin$ **Sepset**$\langle A, B\rangle$.

(ii) orient $A * - * B * - * C$ as $A * - * \underline{B} * - * C$ *iff* $B \in$ **Sepset**$\langle A, B\rangle$.

Step 3. For each pair of variables A and B adjacent in Q', if A and B are d-separated given any subset S of **Possible-d-sepset**$\langle A, B\rangle)\backslash\{A, B\}$ or any subset S of **Possible-d-sepset**$\langle B, A\rangle)\backslash\{A, B\}$ in Q', then remove the edge between A and B, and record $\mathcal{S}$ in **Sepset**$\langle A, B\rangle$ and **Sepset**$\langle B, A\rangle$.

Step 4. Reorient an edge between any pair of variables X and Y ($X \circ\!\!-\!\!\circ Y$) following the orientation rules below.

repeat

  if there is a directed path from A to B, and an edge $A * - * B$ **then**

    orient $A * - * B$ as $A* \rightarrow B$,

  else if B is a collider along $\langle A, B, C\rangle$, B is adjacent to D, and $D \in$ **Sepset**$\langle A, C\rangle$ **then**

    orient $B * - * D$ as $B \leftarrow *D$,

  else if U is a definite discriminating path between A an d B for M in the path $\prod$, and P and R are adjacent to M on U, and $P - M - R$ is a triangle **then**

    if $M \in$ **Sepset**$\langle A, B\rangle$ **then** M is marked as non-collider on subpath $P * - * \underline{M} * - * R$

    else $P * - * M * - * R$ is oriented as $P* \rightarrow M \leftarrow *R$.

    end if

  else if $P* \rightarrow \underline{M} * - * R$ **then**

    orient as $P* \rightarrow M \rightarrow R$.

  end if

until no more edges can be oriented.

Appendix C CCI Details

Algorithm 3 Cyclic Causal Inference (CCI)

Input: A conditional independent oracle for a distribution $\mathcal{P}$, satisfying global directed Markov property and faithfulness conditions w.r.t. a directed graph $\mathcal{G}$ with vertex set $\mathcal{V}$. ($\mathcal{V} = \mathbf{O} \cup \mathbf{L} \cup \mathbf{S}$), where $\mathbf{O}$, $\mathbf{L}$, and $\mathbf{S}$ refer to the sets of observable, latent, and selection variables, respectively.

Output: A PAAG ($\hat{\mathcal{G}}'$) for the Markov equivalence class of MAAGs (Equiv($\mathcal{G}$)).

Step 1. Run FCI's skeleton discovery procedure.

Step 2. Run FCI's collider structure (v-structure) orientation procedure.

Step 3. For any triplet $\langle O_i, O_k, O_j \rangle$, such that we have $O_k \circ\!\!- *O_i$, if $O_i \perp\!\!\!\perp_{\mathcal{G}} O_j \mid \mathbf{Sepset}\langle O_i, O_j \rangle \cup \mathbf{S}$, where $\mathbf{Sepset}\langle O_i, O_j \rangle$ is a separating set discovered in step 1, $O_k \notin \mathbf{Sepset}\langle O_i, O_j \rangle$, $O_i \not\perp\!\!\!\perp_{\mathcal{G}} O_k \mid \mathbf{Sepset}\langle O_i, O_j \rangle \cup \mathbf{S}$ and $O_j \not\perp\!\!\!\perp_{\mathcal{G}} O_k \mid \mathbf{Sepset}\langle O_i, O_j \rangle \cup \mathbf{S}$, then orient $O_k \circ\!\!- *O_i$ as $O_k \leftarrow *O_i$.

Step 4. Find additional non-minimal d-separating sets.

$m = 0$

repeat

  repeat

    Select an ordered triplet $\langle O_i, O_j, O_k \rangle$ with the collider structure $O_i * \!\!\rightarrow O_j \leftarrow *O_k$ such that $|\mathbf{PD\text{-}Sep}(O_i)| \geq m$

    repeat

      Select a subset
      $\mathbf{W} \subseteq \mathbf{PD\text{-}Sep}(O_i) \backslash \{\mathbf{Sepset}\langle O_i, O_k \rangle \cup \{O_j, O_k\}\}$ with m vertices
      $\mathbf{T} = \mathbf{W} \cup \mathbf{Sepset}\langle O_i, O_k \rangle \cup O_j$
      if O_i and O_k are d-separated given $\mathbf{T} \cup \mathbf{S}$, then record the set $\mathbf{T}$ in $\mathbf{Supset}\langle O_i, O_j, O_k \rangle$

    until all subsets
      $\mathbf{W} \subseteq \mathbf{PD\text{-}Sep}(O_i) \backslash \{\mathbf{Sepset}\langle O_i, O_k \rangle \cup \{O_j, O_k\}\}$ have been considered or a d-separating set of O_i and O_k has been recorded in $\mathbf{Supset}\langle O_i, O_j, O_k \rangle$;

  until all triplets $\langle O_i, O_j, O_k \rangle$ with the collider structure $O_i * \!\!\rightarrow O_j \leftarrow *O_k$ and $|\mathbf{PD\text{-}Sep}(O_i)| \geq m$ have been selected;

until all ordered triplets $\langle O_i, O_j, O_k \rangle$ with the collider structure $O_i * \!\!\rightarrow O_j \leftarrow *O_k$ have $|\mathbf{PD\text{-}Sep}(O_i)| < m$;

Step 5. Find all quadruples of vertices $\langle O_i, O_j, O_k, O_l \rangle$ such that O_i and O_k are non-adjacent, $O_i * \!\!\rightarrow O_l \leftarrow *O_k$, and $O_i \perp\!\!\!\perp_{\mathcal{G}} O_k \mid \mathbf{W} \cup \mathbf{S}$ with $O_j \in \mathbf{W}$ and $\mathbf{W} \subseteq \mathbf{O} \backslash \{O_i, O_k\}$. If $O_l \notin \mathbf{W} = \mathbf{Sepset}\langle O_i, O_k \rangle$ as discovered in step 2, then orient $O_j * \!\!\circ\!\!- O_l$ as $O_j * \!\!\rightarrow O_l$. If we also have $O_i * \!\!\rightarrow O_j \leftarrow *O_k$ and $O_l \in \mathbf{W} = \mathbf{Supset}\langle O_i, O_j, O_k \rangle$ as discovered in step 4, then orient $O_j * \!\!\circ\!\!- O_l$ as $O_j * \!\!- O_l$.

15: ***Step 6.*** For any two vertices O_i and O_k, if we have $O_i \perp\!\!\!\perp_{\mathcal{G}} O_k \mid \mathbf{W} \cup \mathbf{S}$ for some $\mathbf{W} \subseteq \mathbf{O} \backslash \{O_i, O_k\}$ discovered in step 1 or step 4 with $O_j \in \mathbf{W}$ but we have $O_i \not\perp\!\!\!\perp_{\mathcal{G}} O_k \mid O_l \cup \mathbf{W} \cup \mathbf{S}$, then orient $O_l \circ\!\!- * O_j$ as $O_l \leftarrow * O_j$.

16: ***Step 7.*** Execute orientation rules until no more endpoints can be oriented. The orientation rules are as follows:

(i) If we have $O_i * \!\!\rightarrow O_j \circ\!\!- * O_k$ with O_i and O_k non-adjacent, then orient $O_j \circ\!\!- * O_k$ as $O_j - * O_k$. Furthermore, if $O_i * \!\!\rightarrow O_j$ is not potentially 2-triangulated w.r.t. O_k, then orient $O_j \circ\!\!\multimap O_k$ as $O_j \rightarrow O_k$.

(ii) If we have $O_i - * O_j \circ\!\!- * O_k$ with O_i and O_k non-adjacent, and $O_j \circ\!\!- O_k$ is not potentially 2-triangulated w.r.t. O_i, then orient $O_j \circ\!\!- * O_k$ as $O_j - * O_k$.

(iii) Suppose we have $O_i * \!\!\rightarrow O_j - O_k$ with O_i and O_k non-adjacent, and $O_i * \!\!\rightarrow O_j$ is potentially 2-triangulated w.r.t. O_k. If $O_i * \!\!\rightarrow O_j$ can be potentially 2-triangulated w.r.t. O_k using only one vertex O_l in the triangle involve $\{O_i, O_j, O_l\}$, then orient $O_i * \!\!\multimap O_l$ as $O_i * \!\!\rightarrow O_l$, $O_j \circ\!\!- * O_l$ as $O_j * \!\!\rightarrow O_l$ and/or $O_j * \!\!\circ\!\!- O_l$ as $O_j * - O_l$. Next, if there exists only one potentially undirected path $\prod_{O_l, O_k}$ between O_l and O_k, then substitute all circle endpoints on $\prod_{O_l, O_k}$ with tail endpoints $(-)$.

(iv) If $O_i * \!\!\rightarrow O_j - * O_k$, there exists a path $\prod = \langle O_k, ..., O_i \rangle$ with at least $n \geq 3$ vertices such that we have $O_h - * O_{h+1}$ for all $1 \leq i \leq n-1$, and we have $O_1 \circ\!\!- * O_n$, then orient $O_1 \circ\!\!- * O_n$ as $O_1 - * O_n$.

(v) If we have the sequence of vertices $\langle O_1, ..., O_n \rangle$ such that $O_i - * O_{i+1}$ with $1 \leq i \leq n-1$, and we have $O_1 \circ\!\!- * O_n$, then orient $O_1 \circ\!\!- * O_n$ as $O_1 - * O_n$.

(vi) If we have $O_k * \!\!\multimap O_i$, there exists a non-potentially 2-triangulated path $\prod = \langle O_i, O_j, O_l, ..., O_k \rangle$ such that $O_k * \!\!\multimap O_i$ is not potentially 2-triangulated w.r.t. O_j, and $O_j * - * O_i * - * O_k$ is an unshielded non-v-structure, then orient $O_k * \!\!\multimap O_i$ as $O_k * - O_i$.

(vii) Suppose we have $O_i \circ\!\!- O_k$, $O_i - * O_k * - O_l$, a non-potentially 2-triangulated path $\prod_1$ from O_i to O_j, and a non-potentially 2-triangulated path $\prod_2$ from O_i to O_l. Let O_m be a vertex adjacent to O_i on $\prod_1$, and let O_n be the vertex adjacent to O_i on $\prod_2$. If further $O_m * - * O_i * - * O_n$ is an unshielded non-v-structure and $O_i \circ\!\!- * O_k$ is not potentially 2-triangulated w.r.t. both O_n and O_m, then orient $O_i \circ\!\!- * O_k$ as $O_i - * O_k$.