Counterfactuals

Kaizhao Liu

February 1, 2024

Contents

1 PDAGs

Definition 1.1 (PDAG)**.** An acyclic partially directed graph, or PDAG for short, is a graph that contains both directed and undirected edges. It is acyclic in the sense that it contains no directed cycles.

Remark. *An acyclic directed graph, or DAG for short, can be regarded as a PDAG without undirected edges.*

Here we list some graph theoretical concepts for a PDAG.

- The *skeleton* of any PDAG is the undirected graph resulting from ignoring the directionality of every edge.
- A *v-structure* in a PDAG $\mathcal{G}$ is an ordered triple of nodes (x, y, z) such that $\mathcal{G}$ contains the directed edges $x \to y$ and $z \to y$, and x and z are not adjacent in $\mathcal{G}$.
- A *clique* in a PDAG is a set of nodes for which every pair of nodes is adjacent.

We also list some notations for a PDAG.

- We use Π_x to denote the set of parents for node x in a PDAG.
- We use N_x to denote the set of neighbors of node x in a PDAG.
- We use $N_{x,y} = N_x \cap N_y$ to denote the set of common neighbors of x and y in a PDAG.
- We use $\Omega_{x,y} = \Pi_x \cap N_y$ to denote the set of parents of node x that are neighbors of node y in a PDAG.

2 MECs

Two network structures are equivalent if the set of distributions that can be represented using one of the structures is identical to the set of distributions that can be represented using the other.

Theorem 2.1. *Two DAGs are Markov equivalent if and only if they have the same skeleton and the same v-structures.*

Proof. □

Definition 2.1 (CPDAG)**.** The CPDAG of a DAG $\mathcal{D}$, denoted as $\mathcal{C}$, is a PDAG that has the same skeleton as $\mathcal{D}$, and an edge is directed in $\mathcal{C}$ if and only if it has the same orientation in every equivalent DAG of $\mathcal{D}$.

We can distinguish two kinds of edges in a DAG according to its CPDAG. A directed edge of a DAG is *compelled* if it occurs in the corresponding CPDAG, otherwise, the directed edge is *reversible*, and the corresponding parents are reversible parents.

3 A Transformational Characterization of MECs

Definition 3.1. A directed edge $x \to y$ is *covered* if $\Pi_y = \Pi_x \cup \{x\}$.

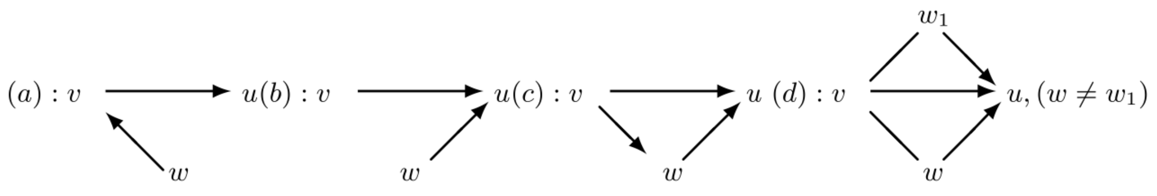
Theorem 3.1. *Let $\mathcal{G}$ be any DAG containing the edge $x \to y$, and let $\mathcal{G}'$ be the directed graph identical to $\mathcal{G}$ except that the edge between x and y in $\mathcal{G}'$ is oriented as $y \to x$. Then $\mathcal{G}'$ is a DAG that is equivalent to $\mathcal{G}$ if and only if $x \to y$ is a covered edge in $\mathcal{G}$.*

4 Characterization of CPDAGs

We introduce some terminologies and the main theorem.

- A PDAG is called a *chain graph* if it contains no semi-directed cycles.
- An undirected graph is *chordal* if every cycle of length greater than or equal to 4 possesses a chord.
- A directed edge $v \to u$ in a PDAG $\mathcal{P}$ is *strongly protected* if $v \to u$ occurs in at least one of the four induced subgraphs of $\mathcal{P}$ in Figure 1.

Figure 1: Configurations where $v \to u$ is strongly protected.



Theorem 4.1. *A graph $\mathcal{C}$ is a CPDAG of a directed acyclic graph $\mathcal{D}$ if and only if $\mathcal{C}$ satisfies the following properties:*

1. *$\mathcal{C}$ is a chain graph;*

2. let $\mathcal{C}_\tau$ be the subgraph induced by τ. $\mathcal{C}_\tau$ is chordal for every chain component τ;

3. $w \to u - v$ does not occur as an induced subgraph of $\mathcal{C}$;

4. every arrow $v \to u$ in $\mathcal{C}$ is strongly protected.

Remark.

The following series of lemma closely inspect a clique of size three in a CPDAG $\mathcal{C}$, where there exists an undirected edge

Lemma 4.1. *The configuration $w \to u - v$ does not occur as an induced subgraph of $\mathcal{C}$.*

Proof. If $w \to u - v$ occurs as an induced subgraph in $\mathcal{C}$, then $w \to u \leftarrow v$ must occurs as a v-structure in some $\mathcal{D} \sim \text{Class}(\mathcal{C})$. By Theorem 2.1, this is impossible. □

Lemma 4.2. *If*
$$\begin{tikzcd} w \arrow[rr] \arrow[dr, dash] & & u \\ & v \arrow[ur, dash] & \end{tikzcd}$$
occurs in $\mathcal{C}$, then there exist $\mathcal{D}_1, \mathcal{D}_2 \in$ Class($\mathcal{C}$) such that
$$\begin{tikzcd} w \arrow[rr] & & u \\ & v \arrow[ul] \arrow[ur] & \end{tikzcd}$$
occurs in $\mathcal{D}_1$ and
$$\begin{tikzcd} w \arrow[rr] \arrow[dr] & & u \arrow[dl] \\ & v & \end{tikzcd}$$
occurs in $\mathcal{D}_2$.

Proof. Any $\mathcal{D} \in \text{Class}(\mathcal{C})$ must contain one of the followings:

$$\begin{tikzcd} \mathrm{w} \arrow[rr] & & \mathrm{u} \\ & \mathrm{v} \arrow[ul] \arrow[ur] & \end{tikzcd}, \quad \begin{tikzcd} \mathrm{w} \arrow[rr] \arrow[dr] & & \mathrm{u} \arrow[dl] \\ & \mathrm{v} & \end{tikzcd}, \quad \begin{tikzcd} \mathrm{w} \arrow[rr] \arrow[dr] & & \mathrm{u} \\ & \mathrm{v} \arrow[ur] & \end{tikzcd}.$$

□

Lemma 4.3. *Let $\{w, u, v\}$ be any three nodes that form a clique of size three in $\mathcal{C}$. If any two of the edges in the clique are undirected, then the third edge is undirected as well.*

Proof. Assume
$$\begin{tikzcd} \mathrm{w} \arrow[rr] \arrow[dr, dash] & & \mathrm{u} \\ & \mathrm{v} \arrow[ur, dash] & \end{tikzcd}$$
occurs in $\mathcal{C}$, we prove by contradiction. According to Lemma 4.2, suppose
$$\begin{tikzcd} \mathrm{w} \arrow[rr] & & \mathrm{u} \\ & \mathrm{v} \arrow[ul] \arrow[ur] & \end{tikzcd}$$
occurs in $\mathcal{D}_1$ and
$$\begin{tikzcd} \mathrm{w} \arrow[rr] \arrow[dr] & & \mathrm{u} \arrow[dl] \\ & \mathrm{v} & \end{tikzcd}$$
occurs in $\mathcal{D}_2$.

As $w \to u$ is compelled, applying Theorem 3.1 to $\mathcal{D}_1$ yields that $w \to u$ is not a covered edge in $\mathcal{D}_1$. Therefore, there exists x such that x is a parent of u but not a parent of w in $\mathcal{G}$. Now consider the subgraph induced by $\{u, v, w, x\}$ in $\mathcal{D}_1$. If there is no edge between x and v, then $x \to u \leftarrow v$ forms a v-structure, which is impossible. So there must be an edge between x and v.

Now consider the subgraph induced by $\{u, v, w, x\}$ in $\mathcal{D}_2$. As $x \to u \leftarrow w$ forms a v-structure in $\mathcal{D}_1$, it also forms a v-structure in $\mathcal{D}_2$ by Theorem 2.1. Moreover, there must be an edge between x and v. If $v \to x$, then $\mathcal{D}_2$ contains a directed cycles. If $x \to v$ in $\mathcal{D}_2$, then $x \to v \leftarrow w$ forms a v-structure, which is also impossible.

□

Lemma 4.4. *Suppose there is a directed path from v to u in $\mathcal{C}$. If there is an edge between v and u, then it must be $v \to u$.*

Proof. Any $\mathcal{D} \in \text{Class}(\mathcal{C})$ must be acyclic. □

Corollary 4.1.1. *If $u - v$ is an undirected edge in $\mathcal{C}$, then $\Pi_u = \Pi_v$.*

Proof. Suppose not. Without loss of generality, let w be any parent of u that is not a parent of v. There are three configuration of the subgraph induced by $\{u, v, w\}$ in general:

$$\begin{tikzcd} \mathrm{w} \arrow[rr] & & \mathrm{u} \arrow[dl, dash] \\ & \mathrm{v} & \end{tikzcd}, \quad \begin{tikzcd} \mathrm{w} \arrow[rr] \arrow[dr, dash] & & \mathrm{u} \\ & \mathrm{v} \arrow[ur, dash] & \end{tikzcd}, \quad \begin{tikzcd} \mathrm{w} \arrow[rr] & & \mathrm{u} \\ & \mathrm{v} \arrow[ul] \arrow[ur, dash] & \end{tikzcd}.$$

The first case is excluded by Lemma 4.1, the second case is excluded by Lemma 4.3, and the third case is excluded by Lemma 4.4. □

Next we focus on the chain graph structure, *i.e.* maximal undirected components in $\mathcal{C}$. A maximal undirected component K of a CPDAG $\mathcal{C}$ is a connected subgraph of $\mathcal{C}$, where each connecting edge is undirected, and for every node $x \in K$, if there is an undirected edge $y - x$ in $\mathcal{C}$, then $y \in K$. For the remainder of this note, we use *undirected component* to mean maximal undirected component.

Lemma 4.5. *For any directed edge $u \to v$ in $\mathcal{C}$, u is a parent of every node reachable by v via undirected edges.*

Proof. Follows by a repeated application of Corollary 4.1.1 along the edges in any undirected path. □

Lemma 4.6. *If there is a directed path of length one or more from u to v in $\mathcal{C}$, then u and v are not in the same undirected component.*

Proof. Suppose not. Consider the last edge $w \to v$ in the directed path. By Lemma 4.5, w is a parent of every node in the undirected component containing v, which means that w is a parent of u. This forms a cycle. □

The next lemma shows that the undirected components of a CPDAG are chordal.

Lemma 4.7. *$\mathcal{C}$ has no undirected chordless k-cycles* $(k \geqslant 4)$.

Proof. If an undirected chordless k-cycle occurs in $\mathcal{C}$, then $\mathcal{C}$ must have at least one v-structure in this cycle. □

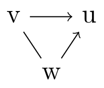
Corollary 4.1.2. *Let u and v be any pair of nodes that are not adjacent in $\mathcal{C}$. Then $N_{x,y}$ is a clique of undirected edges.*

Now we consider the converse of Corollary 4.1.1.

Lemma 4.8. *If $v \to u$ is a directed edge in $\mathcal{C}$, then $\Pi_u \neq \Pi_v \cup \{v\}$.*

Proof. □

Lemma 4.9. *Every directed edge $v \to u$ in $\mathcal{C}$ is strongly protected.*

Proof. According to Lemma, we only need to show that if  occurs in $\mathcal{C}$, then the configuration (d) in Figure 1 occurs as an induced subgraph. □

Now we begin the proof of Theorem

Proof. We first prove the only if direction. □

5 Algorithm:

6 Algorithm:

Definition 6.1. An operator is valid if

- the modified graph of the operator is a PDAG and has a consistent extension;
-

6.1 Compelled and Reversible Edge Insertions

Theorem 6.1. *Let $\mathcal{C}$ be any CPDAG for which nodes x and y are not adjacent. If the result of adding an edge between x and y is a PDAG that admits a consistent extension, then that edge is compelled if and only if $\Pi_x \neq \Pi_y$.*

One reason that Theorem 6.1 is difficult to prove is that after inserting edges into a CPDAG, edges that were reversible before can become compelled, and edges that were compelled before can become reversible. In the following lemmas, we demonstrate a method for determining which edges in a CPDAG will necessarily remain compelled and reversible after an edge addition, based on the topology of the CPDAG.

Lemma 6.1. *Let $\mathcal{C}$ be a CPDAG,and let $\mathcal{P}$ denote the PDAG that results from an edge addition or deletion. If $\mathcal{P}$ admits a consistent extension, then any compelled edge $x \to y$ in $\mathcal{C}$ cannot be compelled in the opposite direction in $\mathcal{P}$.*

Proof. □

6.2 The InsertU Operator

Theorem 6.2. *Let x and y be two nodes that are not adjacent in $\mathcal{C}$. The insertion of the undirected edge $x - y$ is valid if and only if (1) every undirected path between x and y contains a node in $N_{x,y}$, and (2) $\Pi_x = \Pi_y$.*

6.3 The DeleteU Operator

Theorem 6.3. *Let $x - y$ be an undirected edge in CPDAG $\mathcal{C}$. The deletion of $x - y$ is valid if and only if $N_{x,y}$ is a clique of undirected edges.*

6.4 The InsertD Operator

Lemma 6.2. *Let $\mathcal{C}$ be a CPDAG that contains a semi-directed path from x to y. If there exists a directed edge $z \to w$ in this path, then there exists a directed path from z to y in $\mathcal{C}$.*

Theorem 6.4. *Let x and y be any two nodes that are not adjacent in CPDAG $\mathcal{C}$. The insertion of $x \to y$ is valid if and only if (1) every semi-directed path from y to x contains at least one node in $\Omega_{x,y}$, (2) $\Omega_{x,y}$ is a clique of undirected edges, and (3) $\Pi_x \neq \Pi_y$.*

6.5 The DeleteD Operator

Theorem 6.5. *Let $x \to y$ be a directed edge in CPDAG $\mathcal{C}$. The deletion of $x \to y$ is valid if and only if N_y is a clique of undirected edges.*

6.6 The ReverseD Operator

Theorem 6.6. *Let $x \to y$ be a directed edge in CPDAG $\mathcal{C}$. The reversal of $x \to y$ is valid if and only if (1) every semi-directed path from x to y that does not include the edge $x \to y$ contains at least one node in $\Omega_{y,x} \cup N_y$, and (2) $\Omega_{y,x}$ is a clique of undirected edges.*

6.7 The MakeV Operator

Theorem 6.7. *Let $x - z - y$ be any length-two undirected path in CPDAG $\mathcal{C}$ such that x and y are not adjacent. Replacing the undirected edges with directed edges to create the v-structure $x \to z \leftarrow y$ is valid if and only if every undirected path between x and y contains a node in $N_{x,y}$.*

7 Algorithm: Reversible MCMC on MECs

We need to show each operator has a reversible operator.

7.1 The InsertU Operator

Theorem 7.1. *Let x and y be two nodes that are not adjacent in $\mathcal{C}$. The insertion of the undirected edge $x - y$ is reversible if and only if (1) every undirected path between x and y contains a node in $N_{x,y}$, and (2) $\Pi_x = \Pi_y$.*

7.2 The DeleteU Operator

Theorem 7.2. *Let $x - y$ be an undirected edge in CPDAG $\mathcal{C}$. The deletion of $x - y$ is valid if and only if $N_{x,y}$ is a clique of undirected edges.*

7.3 The InsertD Operator

Lemma 7.1. *Let $\mathcal{C}$ be a CPDAG that contains a semi-directed path from x to y. If there exists a directed edge $z \to w$ in this path, then there exists a directed path from z to y in $\mathcal{C}$.*

Theorem 7.3. *Let x and y be any two nodes that are not adjacent in CPDAG $\mathcal{C}$. The insertion of $x \to y$ is valid if and only if (1) every semi-directed path from y to x contains at least one node in $\Omega_{x,y}$, (2) $\Omega_{x,y}$ is a clique of undirected edges, and (3) $\Pi_x \neq \Pi_y$.*

7.4 The DeleteD Operator

Theorem 7.4. *Let $x \to y$ be a directed edge in CPDAG $\mathcal{C}$. The deletion of $x \to y$ is valid if and only if N_y is a clique of undirected edges.*

7.5 The RemoveV Operator

First we show that RemoveC is valid.

Theorem 7.5.

Next we show that RemoveC is reversible.

8 Counting Sizes of MECs

8.1 Size of MEC Determined by the Number of Vertices

Let $\mathcal{U}_{p,n}$ be an undirected and connected chordal graph (UCCG) with p vertices and n edges.