An Introduction to Maximal Ancestral Graphs

Nikolaos Kougioulis

University of Crete
Department of Computer Science

CS-583

November 7, 2023

Definition

A **Bayesian Network (BN)** is a pair (G, Θ) where G is a DAG between random variables $\mathcal{U} = \{X_1, \ldots, X_n\}$ and Θ a set of conditional probabilities $\theta_i = P(X_i|\mathrm{Par}(X_i))\ \forall X_i$. The joint probability on $\mathcal{U}$ decomposes according to the chain rules as

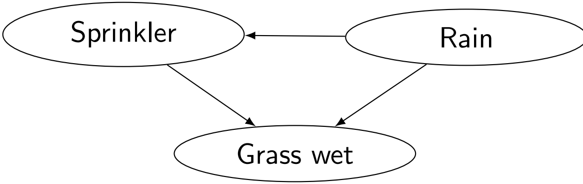
$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\mathrm{Par} X_i)$$

Directed edges

- A directed edge denotes dependence, whereas the absence of an edge denotes (conditional) independence.
- In the causal context, a directed edge $X \rightarrow Y$ denotes that X is a **direct** cause of Y (Y is a **direct** effect of X).

Example (J. Pearl, 2009)

	Sprinkler	
Rain	T	F
F	0.4	0.6
T	0.01	0.99



		Grass wet	
Sprinkler	rain	T	F
F	F	0.4	0.6
F	T	0.01	0.99
T	F	0.01	0.99
T	T	0.01	0.99

Markov Condition

Denote conditional dependence (independence) of two non-empty sets of variables $\mathbf{X}, \mathbf{Y} \subset \mathcal{V}$ given $\mathbf{Z} \subset \mathcal{V}\{\mathbf{X}, \mathbf{Y}\}$ (possibly empty) as $Dep(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$ ($Ind(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$).

Definition

A node is conditionally independent of its nondescendants on the graph, given its parents.

Definition

A node is conditionally independent of its non-effects on the graph, given its direct causes.

Markov Condition - Example

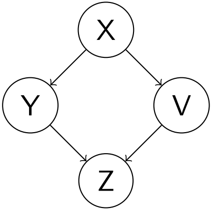


Figure: Figure 1

- $Ind(Y; V|X)$
- $Ind(Z; X|\{Y, V\})$
- $Ind(X; Z|\emptyset)$
- etc.

Colliders and non-colliders

Definition

A node X of a path p in the DAG is called a **collider** if the previous and next nodes of X in the path are into X.

Definition

A node X of a path p in the DAG is called an **unshielded collider** if the previous and next nodes Y, Z of X in the path are into X and Y and Z are not adjacent.

Colliders and non-colliders

Definition

A path p from X to Y is **blocked** by a set of nodes $\mathbf{Z}$ (possibly empty), if some node on p:

- is a collider and neither it or any of its descendants are in $\mathbf{Z}$, or
- is not a collider and is in $\mathbf{Z}$

Definition

If a path p is not blocked by a set of nodes $\mathbf{Z}$, it is said to be **active**.

E.g The path $X - Y - Z$ is blocked by $\mathbf{Y}$, $Y - Z - V$ is blocked by $\mathbf{Z} = \emptyset$ etc.

Definition

Two nodes X and Y are d-separated by $\mathbf{Z}$ iff every path from X to Y is blocked by $\mathbf{Z}$. Otherwise, we say they are **d-connected**.

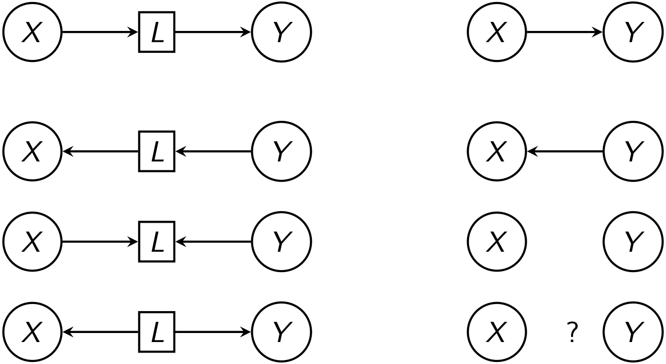
Causal Faithfulness Assumption

Given a causally sufficient set of variables $\mathcal{U}$ in a population N, every conditional independence relation that holds in the density over $\mathcal{U}$ is entailed by the local directed Markov condition for the causal DAG of N.

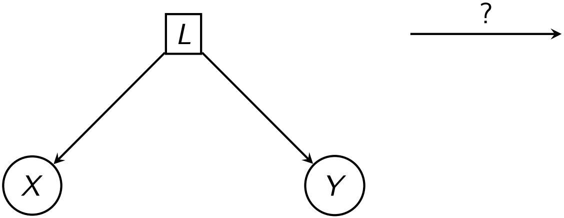
Causal Sufficiency

Assumption that there are unobserved variables.
Unobserved variables are called **latent** or **hidden** variables.

Latent Confounders



Latent Confounders



X and Y are not independent. No edge X and Y represents their independence relation correctly.

Mixed Graphs

Definition

A (directed) **mixed graph** $\mathcal{G}$ is a graph that may contain two kinds of edges: directed edges ($\rightarrow$) and bi-directed edges ($\leftrightarrow$).

- Between any two vertices there is at **most one edge**.
- The two ends of an edge we call **marks**.
- There are two kinds of marks: **arrowhead** ($>$) and **tail** ($-$).
- We say an edge is **into** (**out of**) a vertex if the mark of the edge at the vertex is an arrowhead (or tail).

If

$$\begin{cases} X \leftrightarrow Y \\ X \to Y \\ X \leftarrow Y \end{cases} \text{ in } \mathcal{M} \text{ then } X \text{ is a } \begin{cases} \textbf{spouse} \\ \textbf{parent} \\ \textbf{child} \end{cases} \text{ of } Y \text{ and } \begin{cases} X \in \textbf{sp}(Y) \\ X \in \textbf{pa}(Y) \\ X \in \textbf{ch}(Y) \end{cases}$$

Definition

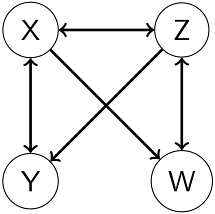
A vertex X is said to be an **ancestor** of a vertex Y, denoted $X \in \textbf{an}(Y)$, if either there exists a directed path $X \to \cdots \to Y$ from X to Y, or $X = Y$.

Ancestral Graphs

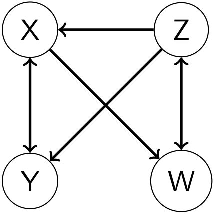
Definition

A mixed (directed) graph is an **ancestral graph** if:

- there are no directed cycles;
- whenever there is an edge $X \leftrightarrow Y$, then there is no directed path from X to Y or from Y to X (no almost directed cycles)



(a) An ancestral graph.



(b) Not an ancestral graph.

Collider Paths in AGs

Definition

- In an ancestral graph, a nonendpoint vertex X on a path is said to be a **collider** if two arrowheads meet at X (i.e., $\rightarrow X \leftarrow, \leftrightarrow X \leftrightarrow, \leftrightarrow X \leftarrow . \rightarrow X \leftrightarrow$).
- All other nonendpoint vertices on a path are called **noncolliders** (i.e., $\rightarrow X \rightarrow, \leftarrow X \leftarrow, \leftarrow X \rightarrow, \leftrightarrow X \rightarrow, \leftarrow X \leftrightarrow$)
- A path along which every nonendpoint is a collider is called a **collider path**.

Definition

In an ancestral graph, a path π between vertices X and Y is active or **m-connecting** relative to a (possibly empty) set of vertices $\mathbf{Z}$, with $X, Y \notin \mathbf{Z}$ if

- every non-collider on π is not a member of $\mathbf{Z}$
- every collider on π is an ancestor of some member of $\mathbf{Z}$

Otherwise we say that **$\mathbf{Z}$ blocks** π.

Example: For the ancestral graph $A \rightarrow B \leftrightarrow C \leftarrow D$:

- The path $\pi_1 = (A, B, C, D)$ is active relative to $\mathbf{Z} = \{B, C\}$
- The path π_1 is not m-connecting relative to $\mathbf{Z} = \emptyset$, $\mathbf{Z} = \{B\}$ or $\mathbf{Z} = \{C\}$.

m-Separation

Definition

- X and Y are said to be **m-separated by Z** if there are no active paths between X and Y relative to $\mathbf{Z}$, i.e if $\mathbf{Z}$ blocks all paths between X and Y.
- Two disjoint sets of variables X and Y are m-separated by $\mathbf{Z}$ if every variable in $\mathbf{X}$ is m-separated from every variable in $\mathbf{Y}$ by $\mathbf{Z}$.

Example: For the ancestral graph $A \to B \leftarrow C \leftarrow D$:

- $\{A\} \perp\!\!\!\perp_m \{D\}$
  $\{A\} \perp\!\!\!\perp_m \{D\} \mid \{B\}$
  $\{A\} \perp\!\!\!\perp_m \{D\} \mid \{C\}$
  since there is no active path relative to $\mathbf{Z} = \emptyset$, $\mathbf{Z} = \{B\}$ and $\mathbf{Z} = \{C\}$ respectively.
- $\{A\} \not\perp\!\!\!\perp_m \{D\} \mid \{B, C\}$ because $\pi_1 = (A, B, C, D)$ is active relative to $\mathbf{Z} = \{B, C\}$

Pairwise Markov property

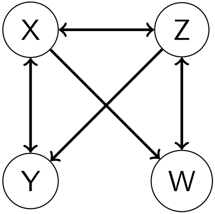
Every missing edge corresponds to a conditional independence relation.

Definition

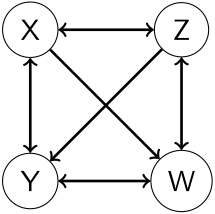
An ancestral graph is **maximal** if for every pair of nonadjacent vertices (a, b) there exists a set $\mathbf{Z}$ with $a, b \notin \mathbf{Z}$ such that a and b are m-separated conditional on $\mathbf{Z}$ (i.e the pairwise Markov property holds).

Definition

An ancestral graph $\mathcal{G}$ is said to be **maximal** if for every pair of non-adjacent vertices (X, Y) there exists a set of $\mathbf{Z}$ $(X, Y \notin \mathbf{Z})$ such that X and Y are m-separated conditional on $\mathbf{Z}$.

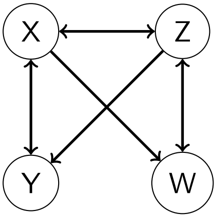


(a) A not maximal ancestral graph.



(b) A maximal ancestral graph.

Maximal Ancestral Graphs



To show that the left AG is not maximal, notice that the only pair of non-adjacent vertices is (Y, W).

- $\mathbf{Z} = \{X\}$, π_1 is *not an active path* since $Z \notin \mathbf{an}(X)$ but $\pi_2 = (Y, Z, W)$ is active since there are no colliders in π_2, Z is a non-collider for π_2 and $Z \notin \mathbf{Z}$
- For $\mathbf{Z} = \{X, Z\}$, $\pi_1 = (Y, X, Z, W)$ is a path that m-connects Y and W.
- For $\mathbf{Z} = \{Z\}$, $\pi_3 = (Y, Z, W)$ is an active path as there are no colliders in π_3, X is a non-collider for π_3 and $X \notin \mathbf{Z}$.
- For $\mathbf{Z} = \emptyset$, π_1 and π_3 are active paths.

Meaning of edges in a MAG

Theorem

A directed edge $X \to Y$ from some node X into another node Y denotes that Y is not an ancestor of X.

Proof.

Let $\mathcal{M}$ be a MAG and X, Y two nodes on $\mathcal{M}$. Assume $X \to Y$ is in $\mathcal{M}$ and Y is an ancestor of X. Then there is a directed path from Y to X, but this means that the MAG $\mathcal{G}$ contains a directed cycle, contradiction. Hence if $X \to Y$, Y is not an ancestor of X. □

Similarly if $X \leftarrow Y$, X is not an ancestor of Y.

With a similar proof as before, we can show that

Theorem

A bidirected edge $X \leftrightarrow Y$ between some nodes X and Y denotes that

- *X is not an ancestor of Y*
- *Y is not an ancestor of X*

Meaning of edges in a MAG

$X \to Y$:

- X is an ancestor of Y
- Y is *not* an ancestor of X
- The above does not rule out **possible latent confounding** between X and Y.

$X \leftrightarrow Y$:

- X is *not* an ancestor of Y
- Y is *not* an ancestor of X
- X and Y are **confounded**.

Maximal Ancestral Graphs

Maximal ancestral graphs (MAGs) are maximal in the sense that **no additional edge may be added to the graph without changing the independence model**.

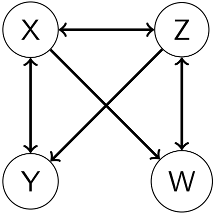
Theorem

If $\mathcal{M} = (V, E)$ is a maximal ancestral graph and $\mathcal{M}$ is a subgraph of $\mathcal{G}^* = (V, E^*)$, then $\mathbf{I}_m(\mathcal{M}) = \mathbf{I}_m(\mathcal{M}^*)$ implies that $\mathcal{M} = \mathcal{M}^*$.

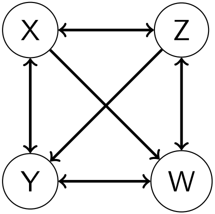
Maximal Ancestral Graphs

Theorem

If $\mathcal{G}$ is an ancestral graph then there exists a **unique** *maximal ancestral graph $\mathcal{M}$ formed by adding $\leftrightarrow$ edges to $\mathcal{G}$ such that* $\mathbf{I}_m(\mathcal{M}) = \mathbf{I}_m(\mathcal{G})$



(a) An ancestral graph $\mathcal{G}$.



(b) The maximal ancestral graph $\mathcal{M}$ from $\mathcal{G}$.

Inducing Paths

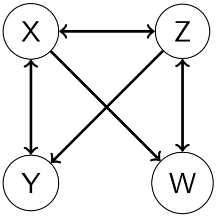
Maximality is closely related to the definition of primitive inducing paths.

Definition

An **inducing path** π **relative to a set L** between two vertices X and Y in an ancestral graph $\mathcal{G}$, is a path on which every non-endpoint vertex, *not* in $\mathbf{L}$ is both a collider on π and an ancestor of *at least one* of the endpoints X and Y.

- Any single-edge path is trivially an inducing path relative to any set of vertices.
- To simplify terminology, we will henceforth refer to inducing paths relative to the empty set simply as **inducing paths**.

Example



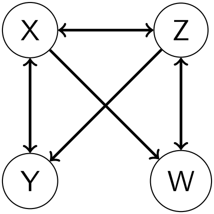
- The path (Y, Z, W) is an inducing path relative to $\{Z\}$ but not an inducing path relative to the empty set (because Z is not a collider)
- The path (Y, X, Z, W) is an inducing path relative to the empty set, because both X and Z are colliders on the path, X is an ancestor of W and Z is an ancestor of Y.

Alternative definition of MAGs

Definition

A mixed graph is called a maximal ancestral graph (MAG) if:

- The graph does not contain any directed or almost directed cycles (ancestral) and
- there is *no inducing path* between any two non-adjacent vertices (maximal)



The ancestral graph is not maximal because the path (Y, X, Z, W) is an inducing path between the non-adjacent vertices Y and W (X and Z are colliders on the path and X, Z ancestors of Z and Y respectively).

A property of MAGs is that they represent the marginal independence models of a DAG over $\mathbf{V} = \mathbf{O} \cup \mathbf{L}$.
This means that given any DAG $\mathcal{D}$ over $\mathbf{V} = \mathbf{O} \cup \mathbf{L}$ there is a MAG $\mathcal{M}$ over $\mathbf{O}$ alone, such that for any disjoint sets $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subset \mathbf{O}$, $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in $\mathcal{D}$ iff they are m-separated by $\mathbf{Z}$ in the MAG $\mathcal{M}$.
This can be cosntructed with the following algorithm:

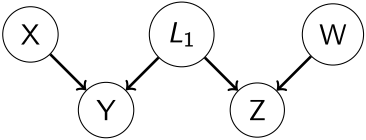
Algorithm DAGs to MAGs

Require: A DAG $\mathcal{D}$ over $\mathbf{O} \cup \mathbf{L}$

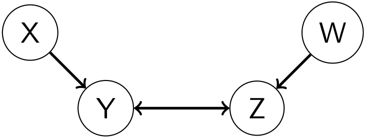
Ensure: A MAG $\mathcal{M}_{\mathcal{D}}$ over $\mathbf{O}$

for all pairs of variables $X, Y \in \mathbf{O}$ **do**
    $\mathcal{M}$ is adjacent to X and Y iff there is an inducing path between them relative to $\mathbf{L}$ in $\mathcal{D}$
end for
for all pairs of adjacent variables X, Y in $\mathcal{M}$ **do**
    if X is an ancestor of Y in $\mathcal{D}$ **then**
        Orient the edge as $X \rightarrow Y$ in $\mathcal{M}$
    else if Y is an ancestor of X in $\mathcal{D}$ **then**
        Orient the edge as $X \leftarrow Y$ in $\mathcal{M}$
    else
        Orient the edge as $X \leftrightarrow Y$ in $\mathcal{M}$
    end if
end for

Example

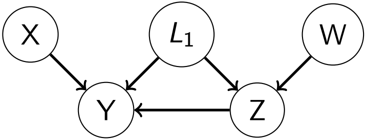


(a) A DAG $\mathcal{D}$ over $\mathbf{O} \cup \mathbf{L}$ where $\mathbf{L} = \{L_1\}$.

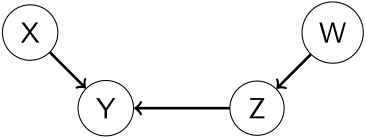


(b) A MAG $\mathcal{D}$ over $\mathbf{O}$ from the DAG $\mathcal{D}$.

Example

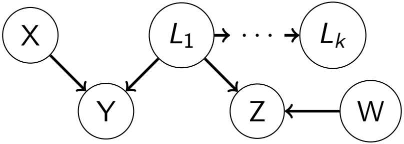


(a) A DAG $\mathcal{D}$ over $\mathbf{O} \cup \mathbf{L}$ where $\mathbf{L} = \{L_1\}$.

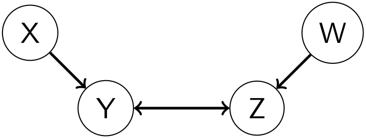


(b) A MAG $\mathcal{D}$ over $\mathbf{O}$ from the DAG $\mathcal{D}$.

Example



(a) A DAG $\mathcal{D}$ over $\mathbf{O} \cup \mathbf{L}$ where $\mathbf{L} = \{L_1, \ldots, L_k\}$.



(b) A MAG $\mathcal{D}$ over $\mathbf{O}$ from the DAG $\mathcal{D}$.

Comments on edges on MAGs

- Notice that as before, if two variables share a hidden common cause and there is a directed edge between them, then the MAG only keeps the directed edge.
- Because MAGs represent ancestral relationships, a directed edge dominates a bi-directed edge.
- There may also exist edges in the MAG that are not present in the underlying causal model

Markov Equivalence

Several MAGs can encode the same conditional independencies via m-separation (as various DAGs can encode the same conditional independencies via d-separation) and are not distinguishable only by correlational patterns.

Definition

Two MAGs $\mathcal{G}_1, \mathcal{G}\_2$ over the same set of vertices are called **Markov equivalent** if for any three disjoint sets of vertices X, Y, Z, X and Y are m-separated by Z in $\mathcal{G}_1$ iff X and Y are m-separated by Z in $\mathcal{G}_2$.

Unshielded colliders

Definition

In a MAG, a path consisting of a triple of vertices X, Y, Z is said to be **unshielded** if X and Z are not adjacent.

Definition

A vertex Y is called an **unshielded collider** if the vertices X, Z are into Y and X and Z are not adjacent.

Discriminating paths

Definition

In a MAG $\mathcal{G}$, a path $\pi = (x, q_1, \ldots, q_p, b, y)$, $p \geq 1$ is called a **discriminating path for** (q_p, b, y) if:

- x is not adjacent to y, and
- every vertex q_i, $1 \leq i \leq p$ is a collider on π and a parent of y.

Markov Equivalence in DAGs

It is known due to Verma and Pearl (1990) that:

Theorem

Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent iff

- *$\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies*
- *$\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders*

Markov Equivalence in MAGs

The following is known as the Spirtes and Richardson Criterion (SRC) (Spirtes and Richardson (1996)):

Theorem

Two MAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent iff

- *$\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies*
- *$\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders and*
- *if π forms a discriminating path for b in $\mathcal{G}_1$ and $\mathcal{G}_2$, then b is a collider on the path π in $\mathcal{G}_1$ if and only if it is a collider on the path π in $\mathcal{G}_2$.*

- Markov Equivalent MAGs form a Markov equivalence class that can be described uniquely by a **partial ancestral graph (PAG)**.
- A PAG $\mathcal{P}$ has the same adjacencies as any MAG in the Markov equivalence class described by $\mathcal{P}$.
- We denote all MAGs in the Markov equivalence class described by a PAG $\mathcal{G}$ by $[\mathcal{G}]$.

Partial Ancestral Graphs

Let *partial mixed graphs* denote the class of graphs containing four types of edges: $\rightarrow$, $\leftarrow$, $\circ\rightarrow$, $\circ$ and three types of end marks: arrowhead ($>$), tail ($-$) and circle $\circ$.

Definition

Let $[\mathcal{M}]$ be the Markov equivalence class of an arbitrary MAG $\mathcal{M}$. The **partial ancestral graph (PAG)** for $[\mathcal{M}]$, $\mathcal{P}_{[\mathcal{M}]}$, is a partial mixed graph such that:

- $\mathcal{P}[\mathcal{M}]$ has the same adjacencies as M (and any member of [M]) does;
- A mark of arrowhead is in $\mathcal{P}[\mathcal{M}]$ if and only if it is shared by all MAGs in $[\mathcal{M}]$;
- A mark of tail is in $\mathcal{P}[\mathcal{M}]$ if and only if it is shared by all MAGs in $[\mathcal{M}]$.

A PAG represents an equivalence class of MAGs by displaying all common edge marks shared by all members in the class and displaying circles for those marks that are not common. This is equivalent to *Partial DAGs (PDAGs)* that represent an equivalence class of DAGs (Spirtes et al. (2000)).

R Ayesha Ali, Thomas S Richardson, and Peter Spirtes. Markov equivalence for ancestral graphs. 2009.

David Maxwell Chickering. A transformational characterization of equivalent bayesian network structures. *UAI'95: Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*, 1995.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Judea Pearl. An introduction to causal inference. *The International Journal of Biostatistics*, 6(2), 2010. URL `https://doi.org/10.2202/1557-4679.1203`.

Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

References II

Thomas S Richardson and Peter Spirtes. Causal inference via ancestral graph models. *Oxford Statistical Science Series*, pages 83–105, 2003.

Richard Scheines. An introduction to causal inference. 1997.

Peter Spirtes. An anytime algorithm for causal inference. In *International Workshop on Artificial Intelligence and Statistics*, pages 278–285. PMLR, 2001.

Peter Spirtes and Thomas Richardson. A polynomial time algorithm for determining dag equivalence in the presence of latent variables and selection bias. In *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics*, pages 489–500. Citeseer, 1996.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

References III

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270, 1990.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008.