Bayesian Networks

Chapter 14

Bayes Nets (so far)

- A belief network is automatically acyclic by construction.
- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
- The parents of a node n are those variables on which n directly depends.
- A belief network is a graphical representation of dependence and independence:
  - A variable is independent of its non-descendants given its parents.

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

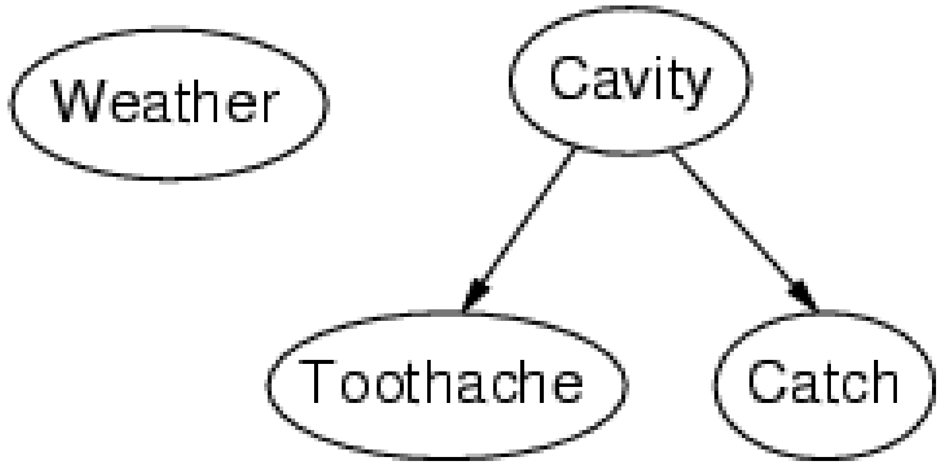
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}\,(X_i \mid \text{Parents}\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

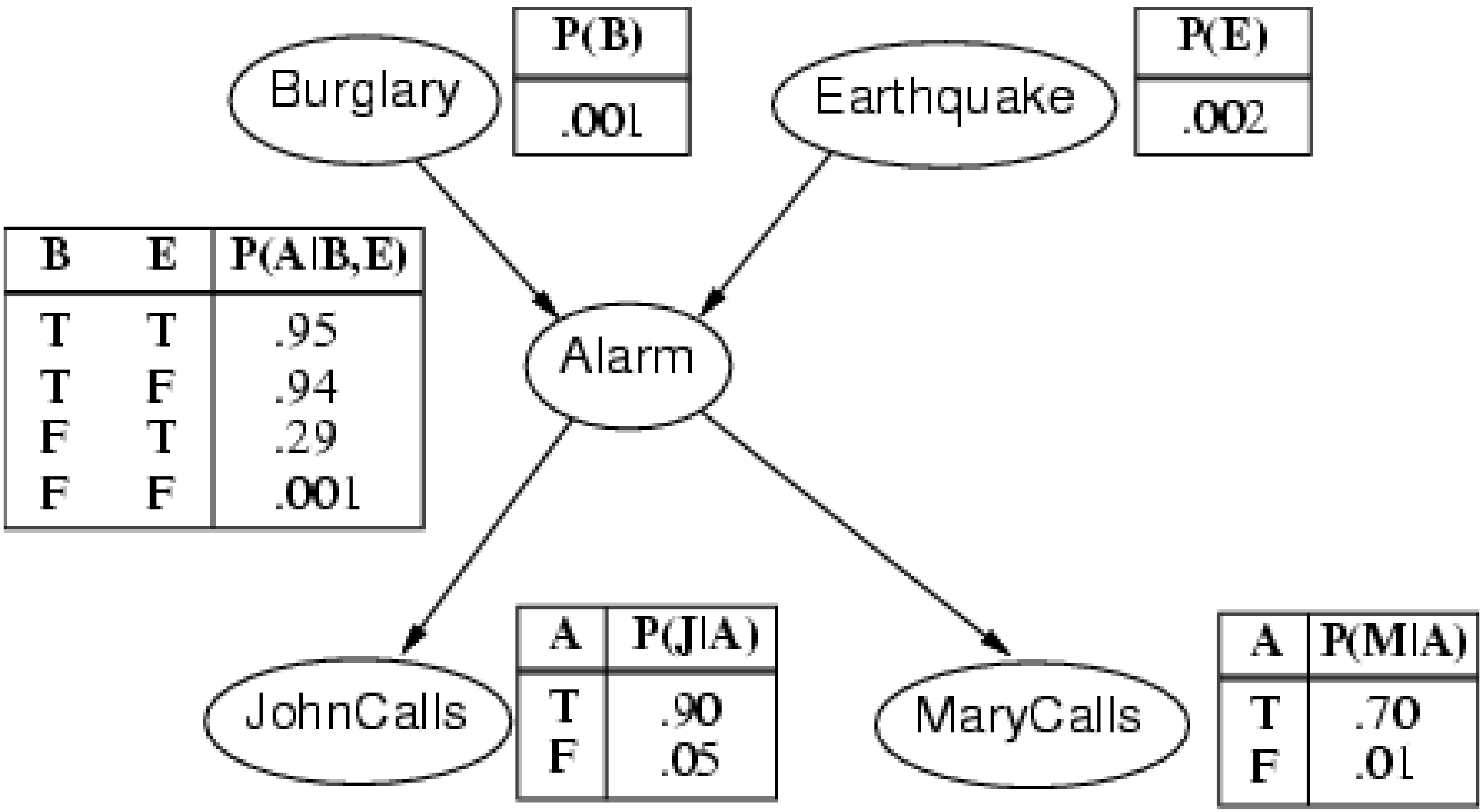
Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Example contd.



P(B)
.001

P(E)
.002

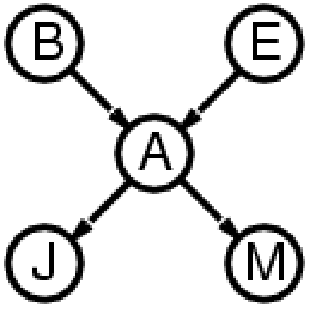
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for $X_i = true$ (the number for $X_i = false$ is just $1-p$)

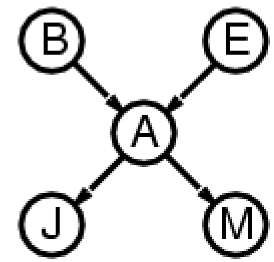
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers

- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)

Semantics

The full joint distribution is defined as the product of the local conditional distributions:



$$\boldsymbol{P}\,(X_1, \ldots ,X_n) = \pi_{i\,=\,1}^{n}\, \boldsymbol{P}\,(X_i \mid Parents(X_i))$$

e.g., $\boldsymbol{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= \boldsymbol{P}\,(j \mid a)\, \boldsymbol{P}\,(m \mid a)\, \boldsymbol{P}\,(a \mid \neg b, \neg e)\, \boldsymbol{P}\,(\neg b)\, \boldsymbol{P}\,(\neg e)$$

Constructing Bayesian networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For i = 1 to n
   - add X_i to the network
   - select parents from $X_1, \ldots, X_{i-1}$ such that

$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

n

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$$

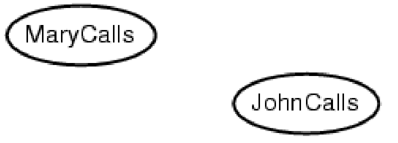
(chain rule)

$$= \pi_{i=1} \mathbf{P}(X_i \mid Parents(X_i))$$

(by construction)

Example

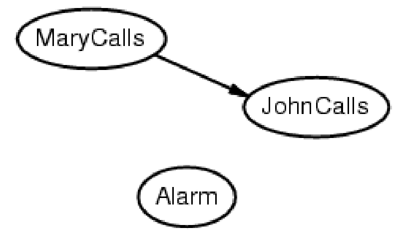
- Suppose we choose the ordering *M, J, A, B, E*



P*(J | M) =* ***P****(J)?*

Example

- Suppose we choose the ordering *M, J, A, B, E*



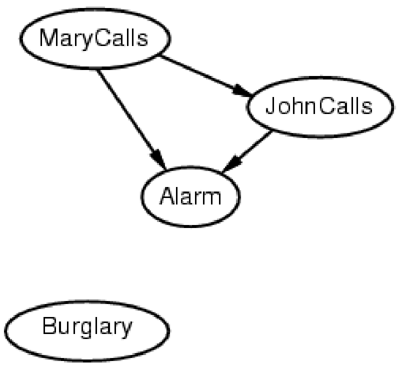
$\boldsymbol{P}(J \mid M) = \boldsymbol{P}(J)$?

No

$\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A \mid J)$? $\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A)$?

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

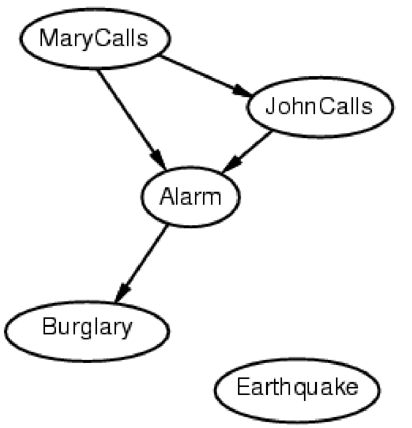
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$?

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

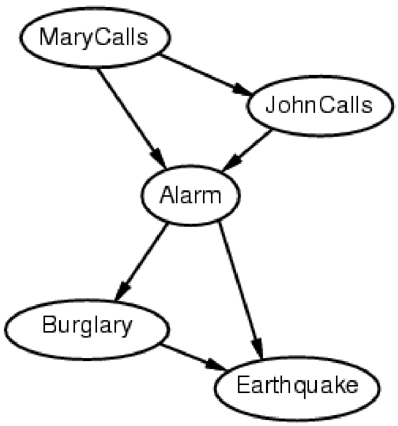
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$?

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

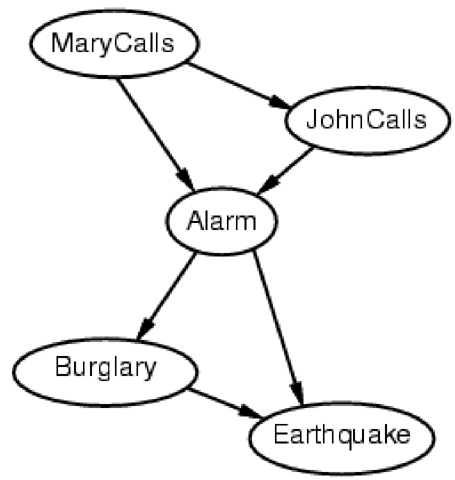
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$? **No**

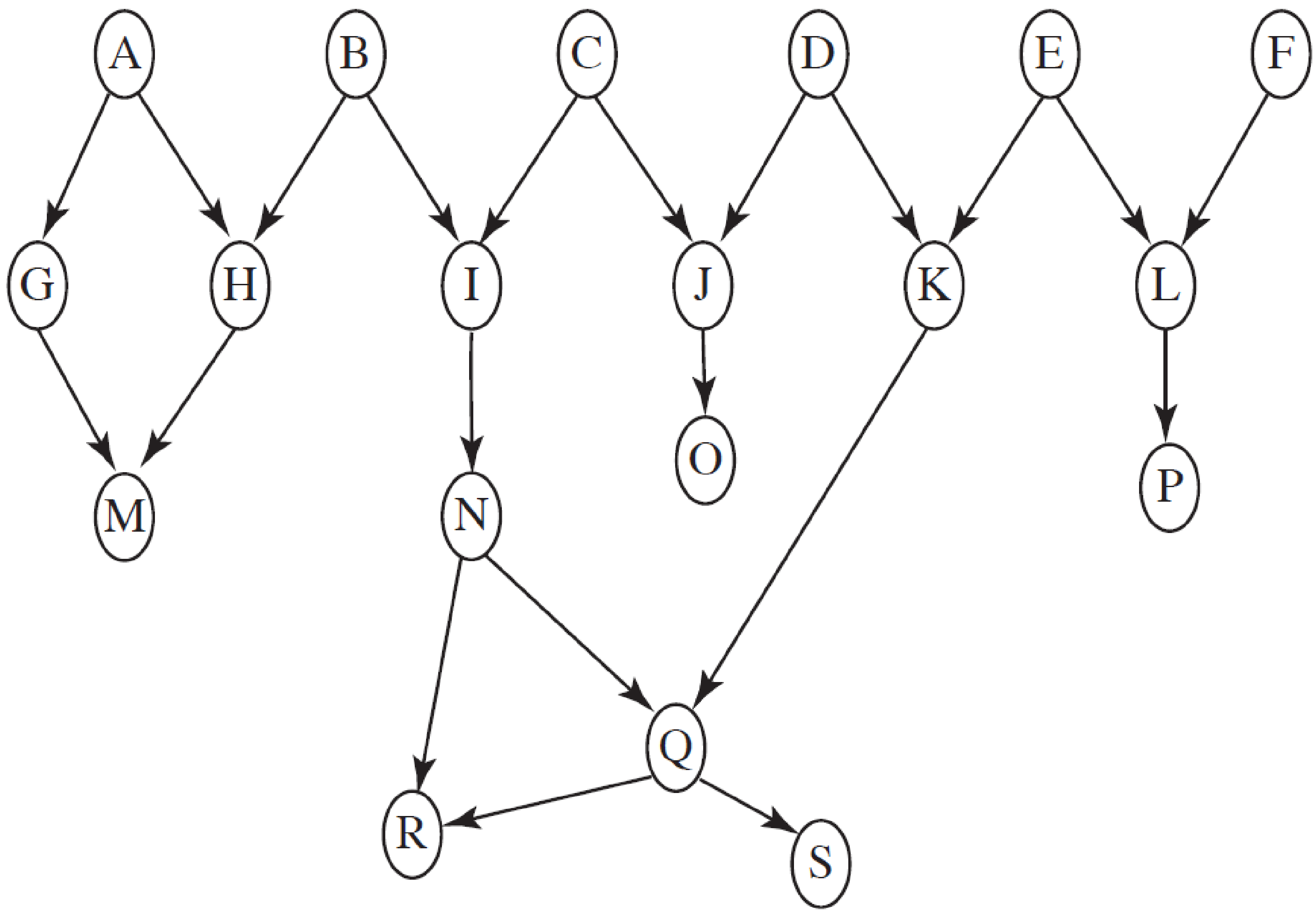
$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$? **Yes**

Example contd.



- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

More on Conditional Independence - Example



Conditional Independence

- On which given probabilities does P(N) depend?
- If you were to observe a value for B, which variables‘ probabilities will change?
- If you were to observe a value for N, which variables'

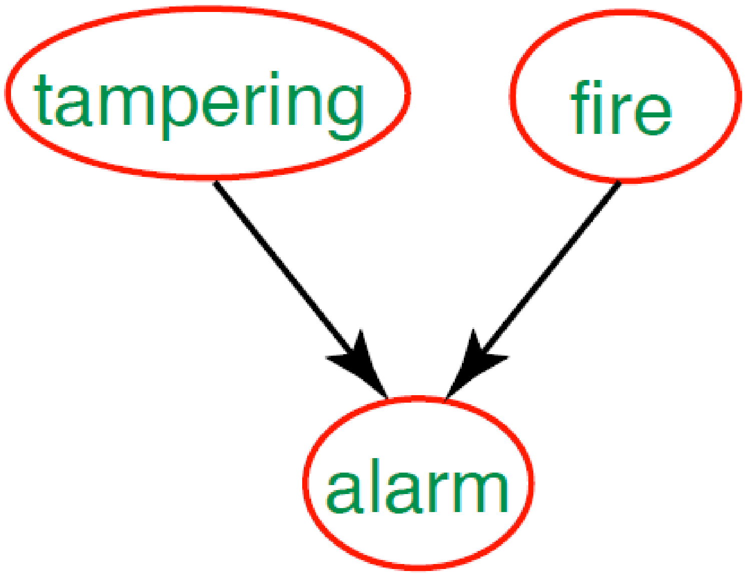
probabilities will change?

- Suppose you had observed a value for M; if you were to then observe a value for N, which variables' probabilities will change?
- Suppose you had observed B and Q; which variables‘ probabilities will change when you observe N?

Conditional Independence

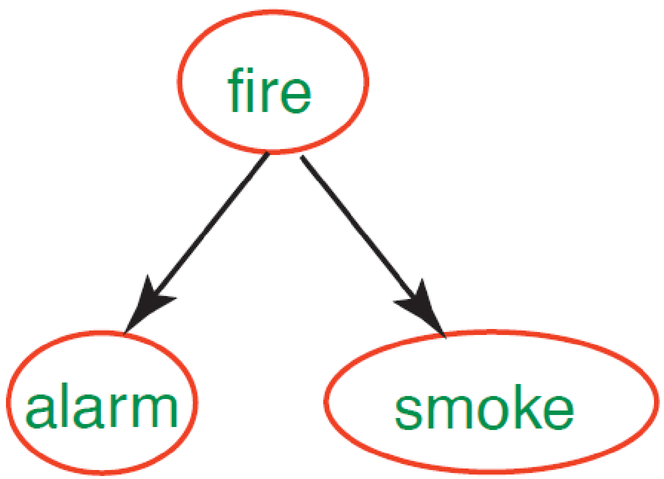
- If you observe variable Y , the variables whose posterior probability is different from their prior are:
  - The ancestors of Y and their descendants.
- Intuitively (if you have a causal belief network):
  - You do abduction to possible causes and prediction from the causes.
- **<u>Three important</u>** cases to consider

Case 1: Common Descendant



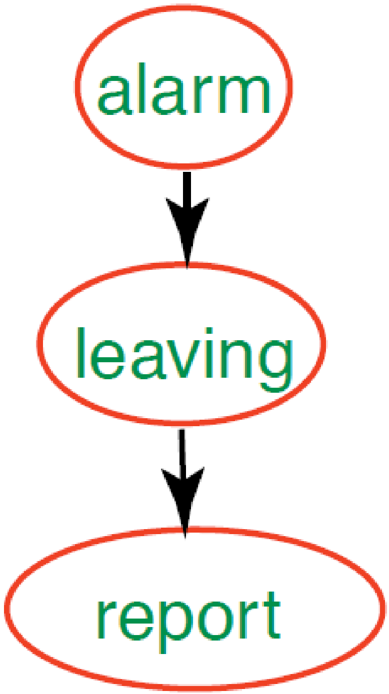
- tampering and fire are independent
- tampering and fire are dependent given alarm
- Intuitively, tampering can <u>explain away</u> fire

Case 2: Common Ancestor



- alarm and smoke are dependent
- alarm and smoke are independent given fire
- Intuitively, fire can explain alarm and smoke; learning one can affect the other by changing your belief in fire

Case 3: Chain



- alarm and report are dependent
- alarm and report are independent given leaving
- Intuitively, the only way that the alarm affects report is by affecting leaving .

Pruning Irrelevant Variables

- Suppose you want to compute $P(X_j | e_1 \ldots e_k)$:
  - Prune any variables that have no observed or queried descendents.
  - Connect the parents of any observed variable.
  - Remove arc directions.
  - Remove observed variables.
  - Remove any variables not connected to X in the resulting (undirected) graph.