Bayesian Networks

Lecture Outline

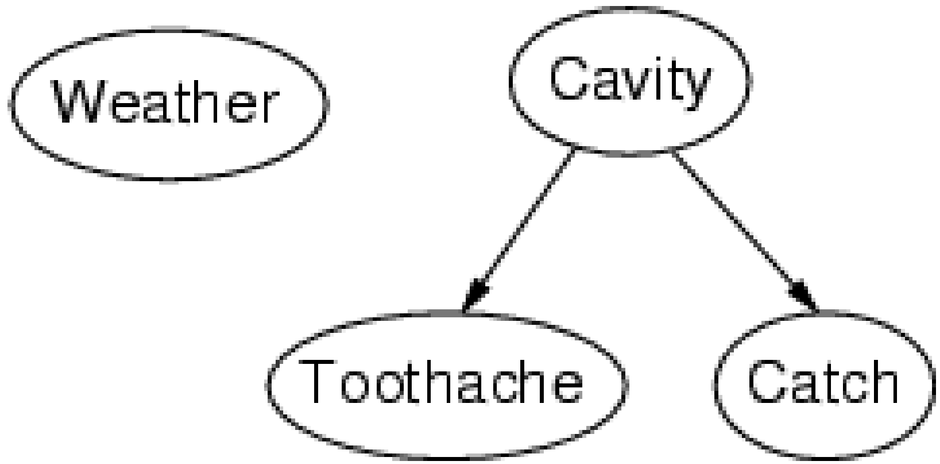
1. Introduction to Bayesian networks
2. How to use Bayesian network
3. How to construct Bayesian network

Bayesian networks

- A Bayesian network is a graph with the following:
  - 1. Nodes: Set of random variables
  - 2. Directed links: The intuitive meaning of a link from node X to node Y is that X has a direct influence on Y
  - 3. Each node has a conditional probability table (CPT) that quantifies the effects that the parent have on the node.
  - 4. The graph has no directed cycles (DAG)

Example

- Topology of network encodes conditional independence assertions:



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

Example

- Burglar alarm at home
  - Turns on if a burglary has occurred.
  - Fairly reliable at detecting a burglary but responds at times, to minor earthquakes as well.
- Two neighbors John and Mary may call the police on hearing alarm.
  - John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
  - Mary likes loud music and sometimes misses the alarm altogether

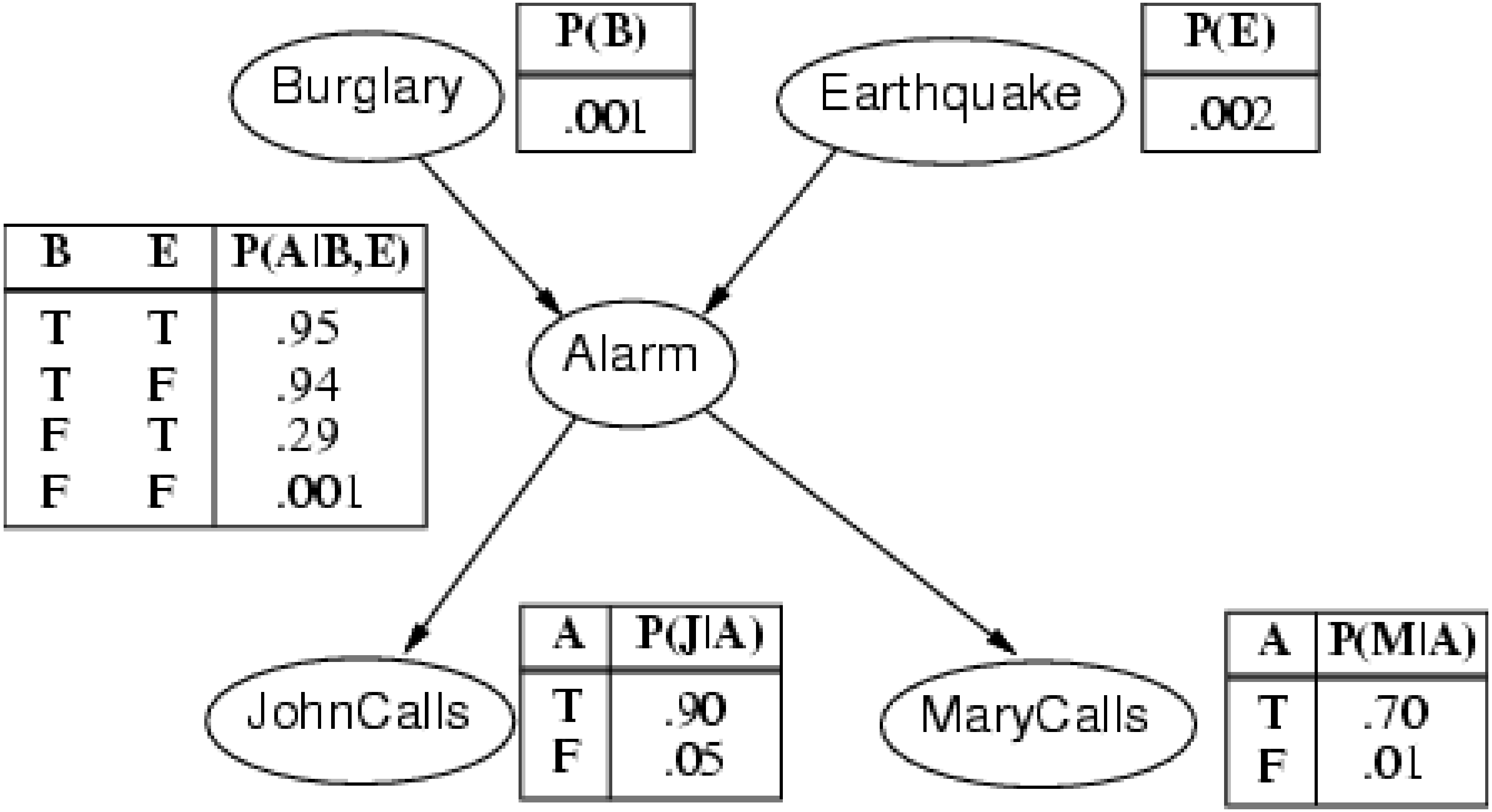
Example

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

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

- Deal with questions like: neighbor John calls, but neighbor Mary doesn't call. Is there a burglar?

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

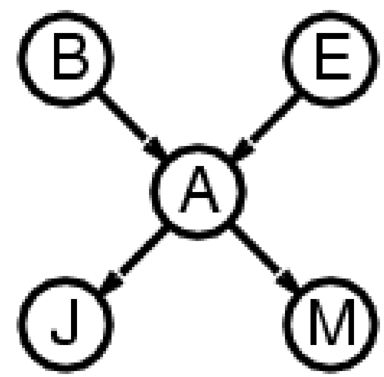
A	P(M|A)
T	.70
F	.01

Full joint probability distribution

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

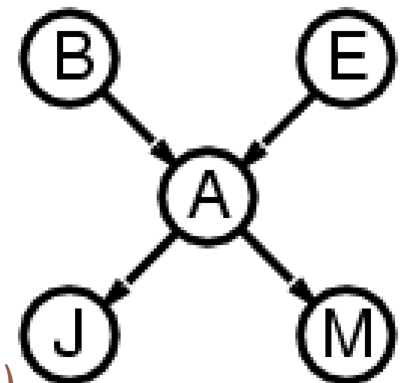
$$\boldsymbol{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \boldsymbol{P}(X_i \mid X_1, \ldots, X_{i-1}) \text{ (chain rule)}$$

$$= \pi_{i=1}^{n} \boldsymbol{P}(X_i \mid Parents(X_i)) \text{ (local conditional)}$$



Full joint probability distribution

- Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:



$\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

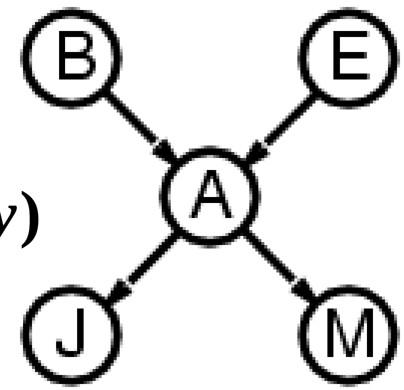
$= \mathbf{P}(j \mid a)\ \mathbf{P}(m \mid a)\ \mathbf{P}(a \mid \neg b, \neg e)\ \mathbf{P}(\neg b)\ \mathbf{P}(\neg e)$

$= 0.9 * 0.70 * 0.01 * 0.999 * 0.998$

$= 0.00063$

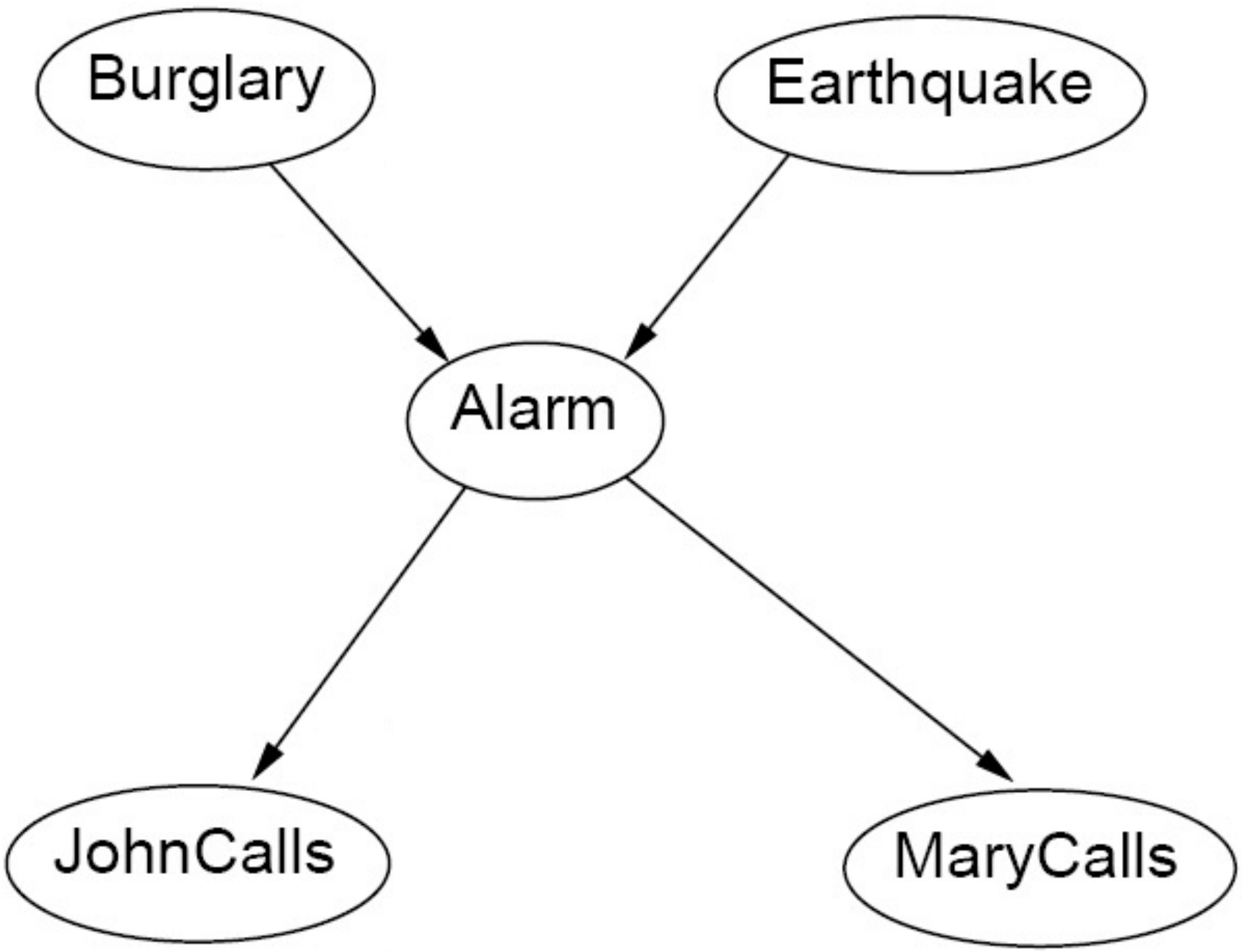
Answering a Query: Inference by Enumeration

- Suppose we want to find *P(X/**e**)*, where *X* is a set of query variables and **e** is the evidence we have
  - E.g. Probability that burglary occurred given both John and Mary had called. Ie. *P(B=true / J=true, M=true)* written as *P(b / j, m)* for short.

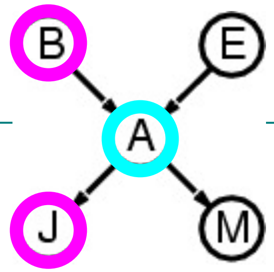


$$P(X/e) = \frac{p(X,e)}{P(e)} = \frac{\sum_y P(X,e,y)}{\sum_x \sum_y P(x,e,y)} = \alpha \sum_y P(X,e,y)$$

Example: Burglar Alarm



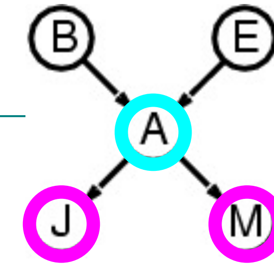
Conditional independence relationships



- Suppose the alarm went off. Does knowing whether there was a burglary change the probability of John calling?

  P(John | Alarm, Burglary) = P(John | Alarm)

Conditional independence relationships



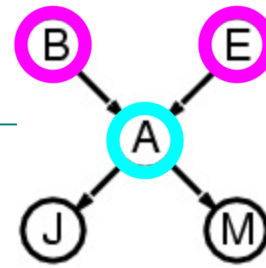
- Suppose the alarm went off. Does knowing whether there was a burglary change the probability of John calling?

  P(John | Alarm, Burglary) = P(John | Alarm)

- Suppose the alarm went off. Does knowing whether John called change the probability of Mary calling?

  P(Mary | Alarm, John) = P(Mary | Alarm)

Conditional independence relationships



- Suppose the alarm went off. Does knowing whether there was a burglary change the probability of John calling?

  P(John | Alarm, Burglary) = P(John | Alarm)

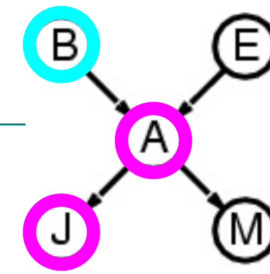
- Suppose the alarm went off. Does knowing whether John called change the probability of Mary calling?

  P(Mary | Alarm, John) = P(Mary | Alarm)

- Suppose the alarm went off. Does knowing whether there was an earthquake change the probability of burglary?

  P(Burglary | Alarm, Earthquake) != P(Burglary | Alarm)

Conditional independence relationships



- Suppose the alarm went off. Does knowing whether there was a burglary change the probability of John calling?

  P(John | Alarm, Burglary) = P(John | Alarm)

- Suppose the alarm went off. Does knowing whether John called change the probability of Mary calling?

  P(Mary | Alarm, John) = P(Mary | Alarm)

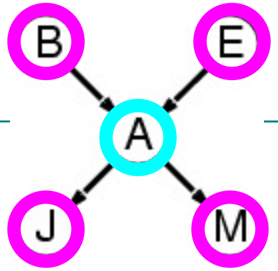
- Suppose the alarm went off. Does knowing whether there was an earthquake change the probability of burglary?

  P(Burglary | Alarm, Earthquake) != P(Burglary | Alarm)

- Suppose there was a burglary. Does knowing whether John called change the probability that the alarm went off?

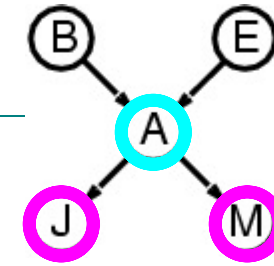
  P(Alarm | Burglary, John) != P(Alarm | Burglary)

Conditional independence relationships



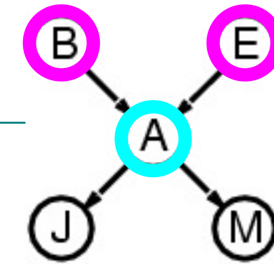
- John and Mary are conditionally independent of Burglary and Earthquake given Alarm
  - *Children* are conditionally independent of *ancestors* given *parents*

Conditional independence relationships



- John and Mary are conditionally independent of Burglary and Earthquake given Alarm
  - *Children* are conditionally independent of *ancestors* given *parents*
- John and Mary are conditionally independent of each other given Alarm
  - *Siblings* are conditionally independent of each other given *parents*

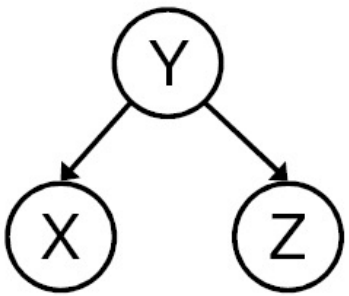
Conditional independence relationships



- John and Mary are conditionally independent of Burglary and Earthquake given Alarm
  - *Children* are conditionally independent of *ancestors* given *parents*
- John and Mary are conditionally independent of each other given Alarm
  - *Siblings* are conditionally independent of each other given *parents*
- Burglary and Earthquake are *not* conditionally independent of each other given Alarm
  - *Parents* are *not* conditionally independent given *children*

Conditional independence

- Common cause



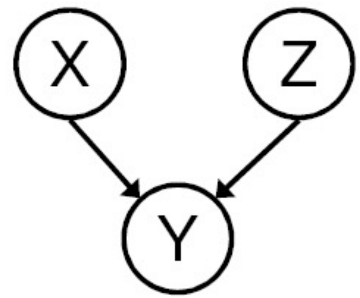
Y: Project due

X: Newsgroup busy

Z: Lab full

- Are X and Z independent?
  - No
- Are they conditionally independent given Y?
  - Yes

- Common effect



X: Raining

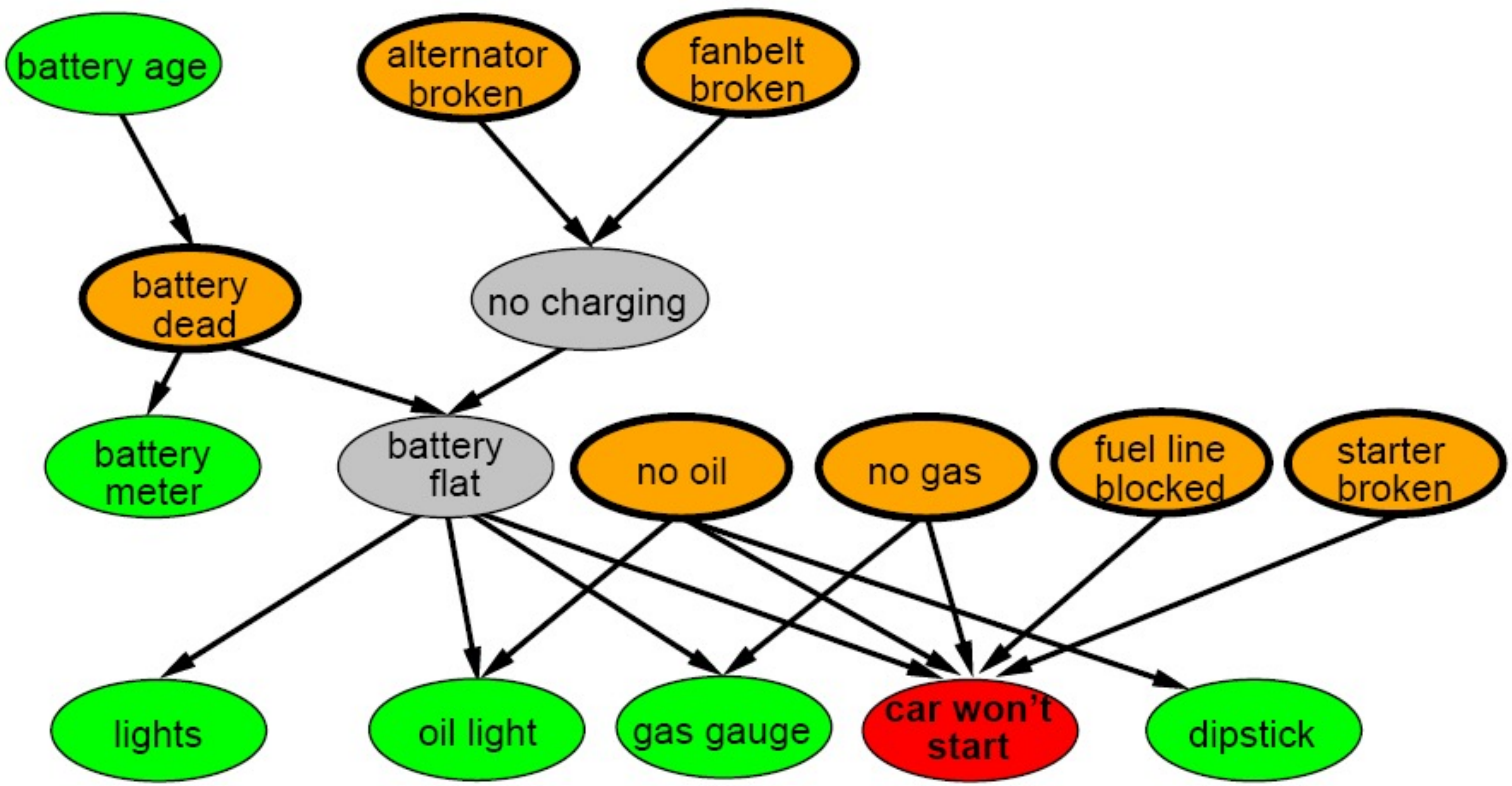
Z: Ballgame

Y: Traffic

- Are X and Z independent?
  - Yes
- Are they conditionally independent given Y?
  - No

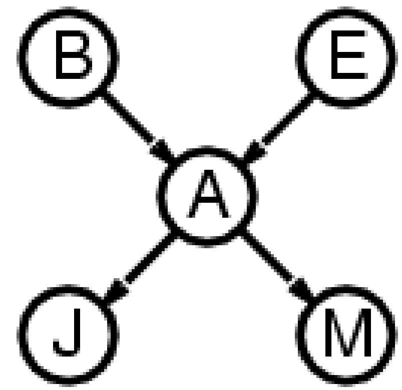
A more realistic Bayes Network: Car diagnosis

- Initial observation: car won’t start
- Orange: “broken, so fix it” nodes
- Green: testable evidence
- Gray: “hidden variables” to ensure sparse structure, reduce parameteres



Compactness

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



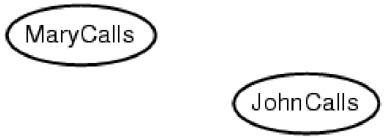
- Each row requires one number p for X_i = *true* (the number for X_i = *false* is just $1-p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. $2^5-1 = 31$)

Constructing Bayesian networks

- 1. Choose the set of relevant variables X that describe the domain
- 2. Choose an ordering for the variables (important step)
  - Any ordering will work but when cause precedes effect, the network becomes more compact
- 3. While there are variables left:
  - a) Pick a variable X_i and add a node for it
  - b) Set Parents(X_i) to some minimal set of existing nodes such that the conditional independence property is satisfied
  - c) Define the conditional probability table for X_i.

Example

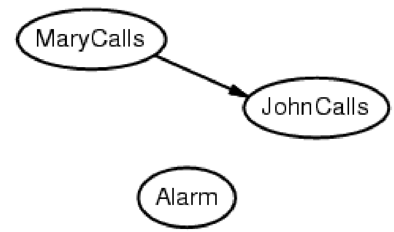
- Suppose we choose the ordering *M*, *J*, *A*, *B*, *E*
-



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

Example

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



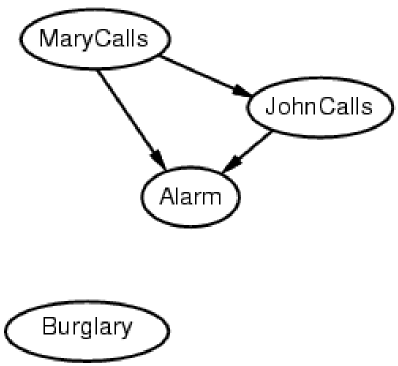
$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

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

Example

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



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

No

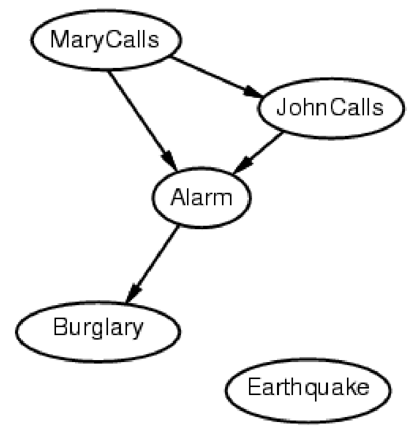
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

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

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

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**

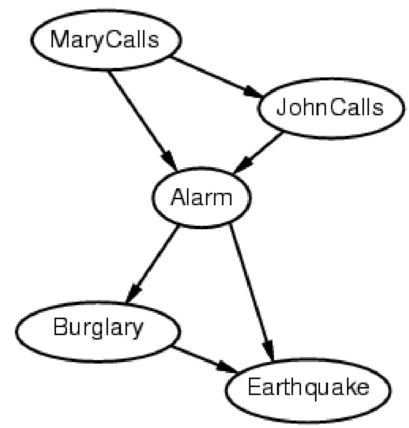
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

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

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

Example

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



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

No

***P**(A | J, M) = **P**(A | J)? **P**(A | J, M) = **P**(A)?* **No**

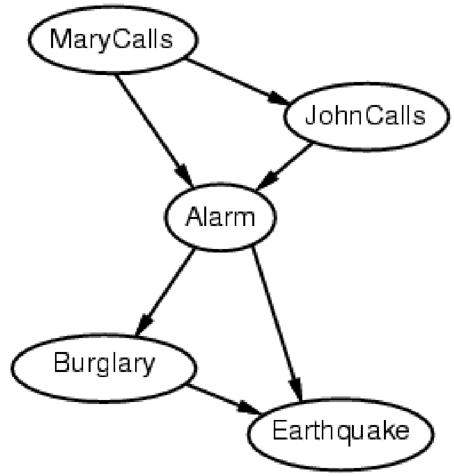
***P**(B | A, J, M) = **P**(B | A)?* **Yes**

***P**(B | A, J, M) = **P**(B)?* **No**

***P**(E | B, A ,J, M) = **P**(E | A)?* **No**

***P**(E | B, A, J, M) = **P**(E | A, B)?* **Yes**

Example contd.



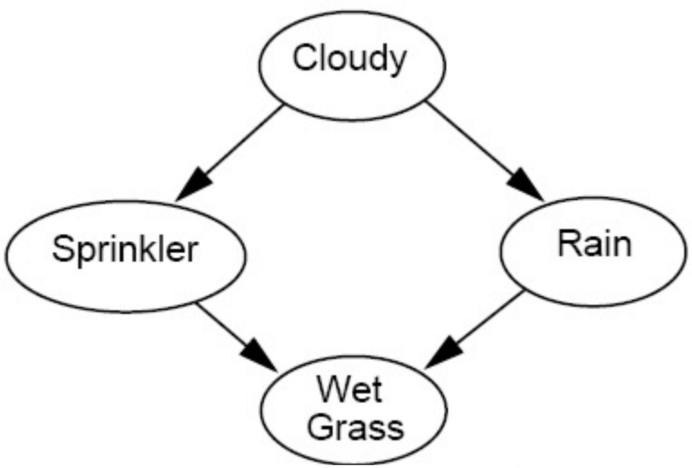
- This network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

- Deciding conditional independence is hard in noncausal directions

  (Causal models and conditional independence seem hardwired for humans!)
- A causal model (links from cause to effects) usually is more compact than a diagnostic (links from effects to causes) model.
- So it's better to order variables as, root causes first, then variables they influence and so on…

Another example

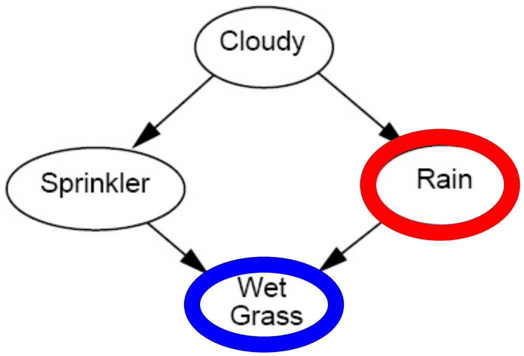
- Variables: *Cloudy, Sprinkler, Rain, WetGrass*



Another example

- Given that the grass is wet, what is the probability that it has rained?

$$P(r \mid w)$$



Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct