Another look at Bayesian inference

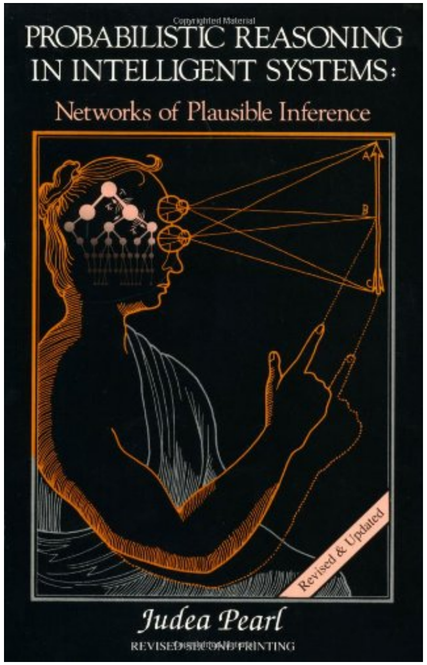
- A general scenario:
  - Query *variables:* **X**
  - *Evidence* (*observed*) variables and their values: **E** = **e**
  - *Unobserved* variables: **Y**
- **Inference problem**: answer questions about the query variables given the evidence variables
  - This can be done using the posterior distribution P(**X** | **E** = **e**)
  - In turn, the posterior needs to be derived from the full joint P(**X**, **E**, **Y**)

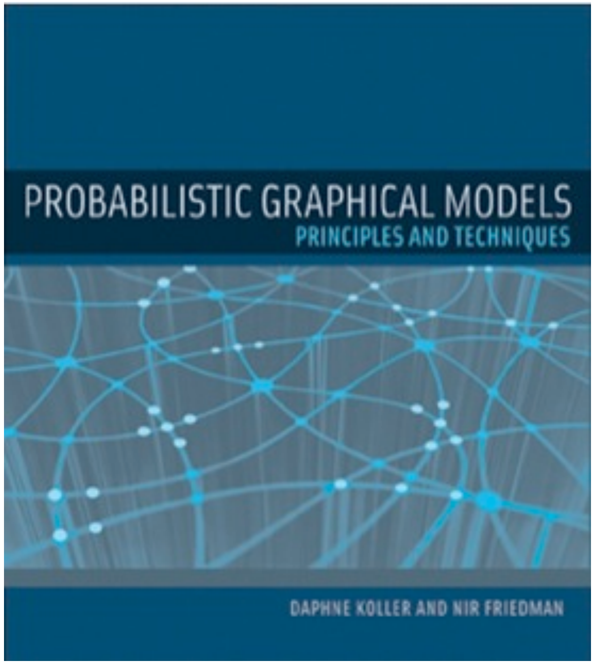
$$P(X \mid E = e) = \frac{P(X, e)}{P(e)} \propto \sum_{y} P(X, e, y)$$

- Bayesian networks are a tool for representing joint probability distributions efficiently

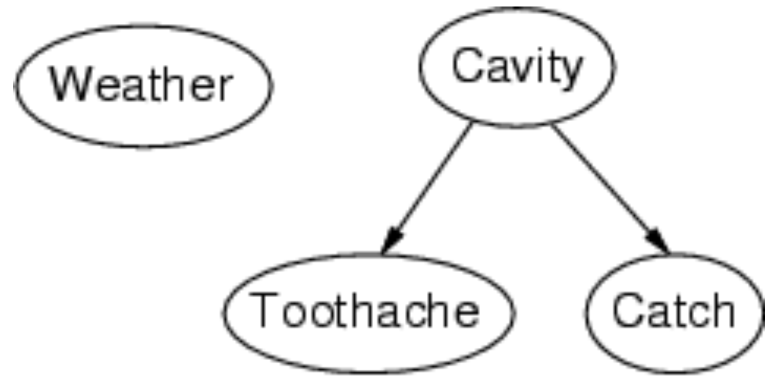
Bayesian networks

- More commonly called *graphical models*
- A way to depict conditional independence relationships between random variables
- A compact specification of full joint distributions





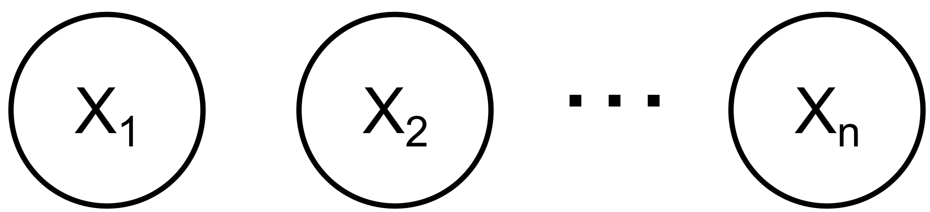
Bayesian networks: Structure



- **Nodes:** random variables
- **Arcs:** interactions
  - An arrow from one variable to another indicates direct influence
  - Must form a directed, *acyclic* graph

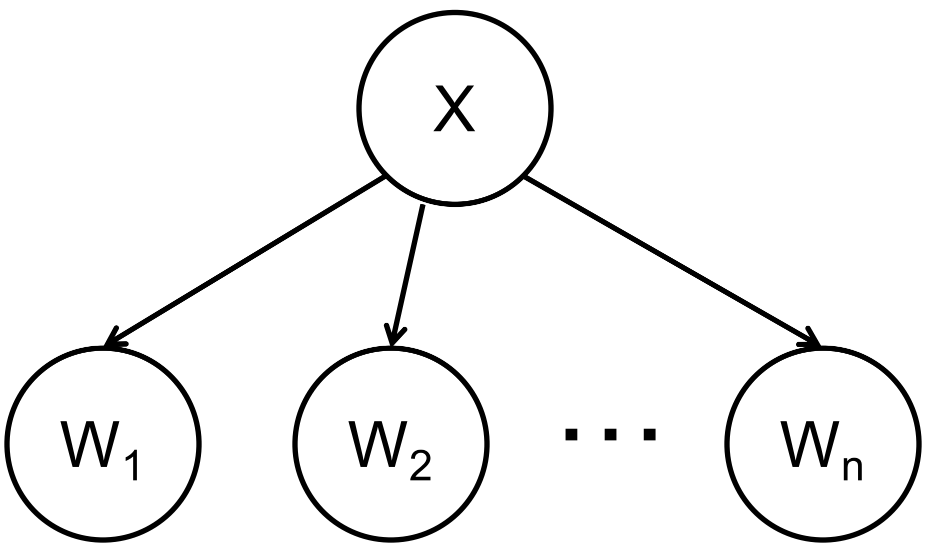
Example: N independent coin flips

- Complete independence: no interactions



Example: Naïve Bayes document model

- Random variables:
  - X: document class
  - $W_1, \ldots, W_n$: words in the document



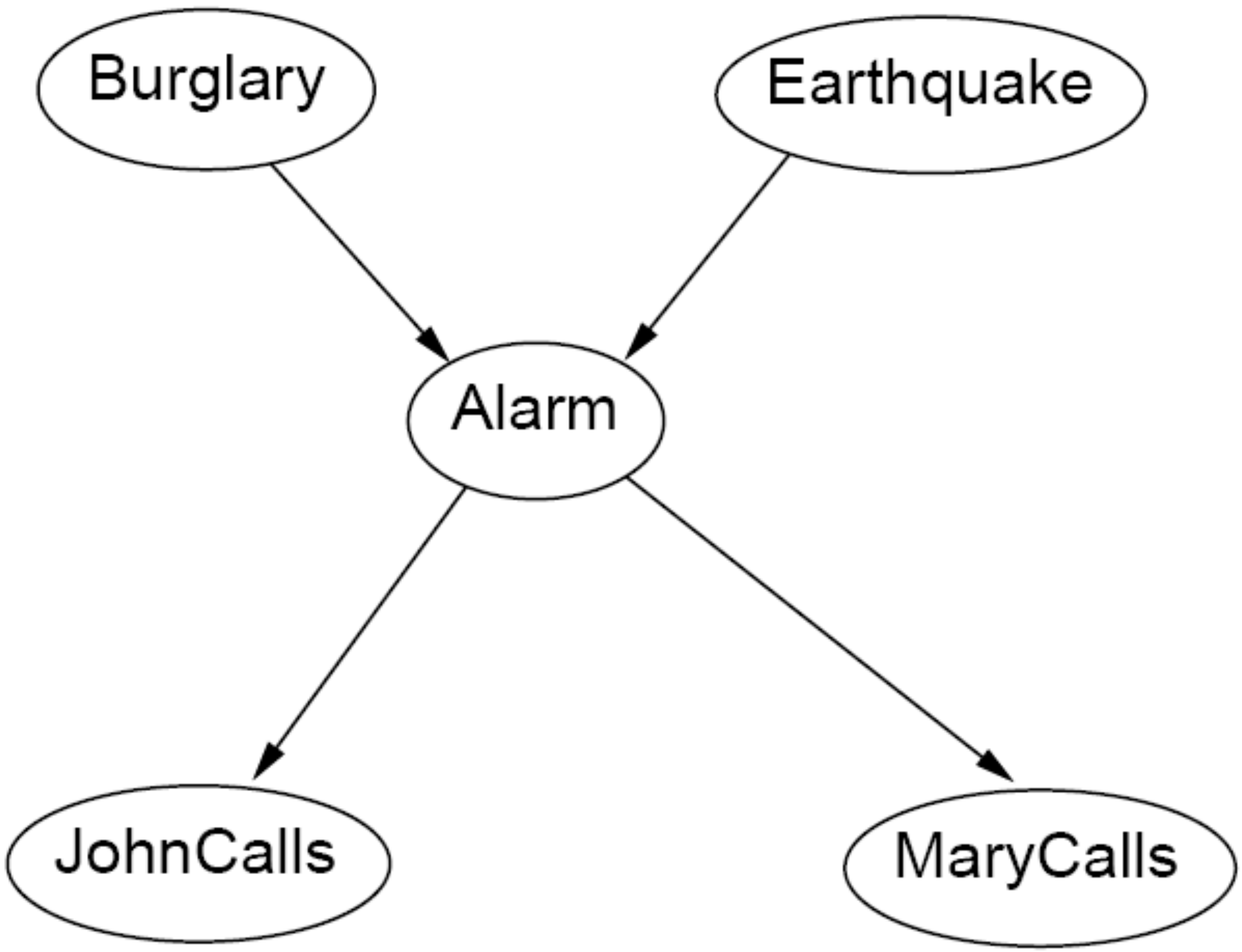
Example: Burglar Alarm

- I have a burglar alarm that is sometimes set off by minor earthquakes. My two neighbors, John and Mary, promised to call me at work if they hear the alarm
- Example inference tasks
  - Suppose Mary calls and John doesn’t call. What is the probability of a burglary?
  - Suppose there is a burglary and no earthquake. What is the probability of John calling?
  - Suppose the alarm went off. What is the probability of burglary?
  - …

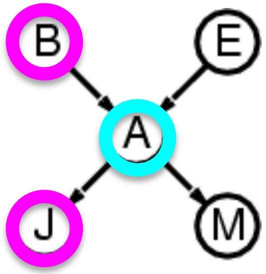
Example: Burglar Alarm

- I have a burglar alarm that is sometimes set off by minor earthquakes. My two neighbors, John and Mary, promised to call me at work if they hear the alarm
- What are the random variables?
  – Burglary, Earthquake, Alarm, John, Mary
- What are the direct influence relationships?
  – 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: Burglar Alarm



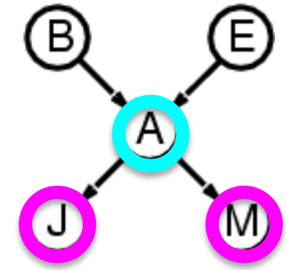
Conditional independence relationships



- 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



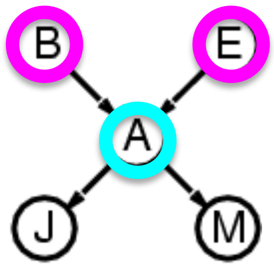
- Suppose the alarm went off. Does knowing whether there was a burglary change the probability of John calling?

  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)

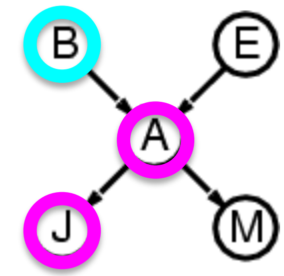
- Suppose the alarm went off. Does knowing whether John called change the probability of Mary calling?

  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)

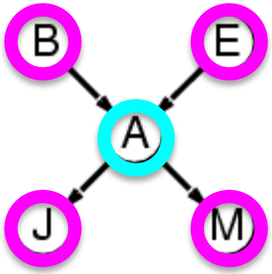
- Suppose the alarm went off. Does knowing whether there was an earthquake change the probability of burglary?

  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?

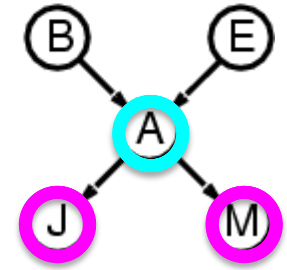
  P(Alarm | Burglary, John) != P(Alarm | Burglary)

Conditional independence relationships



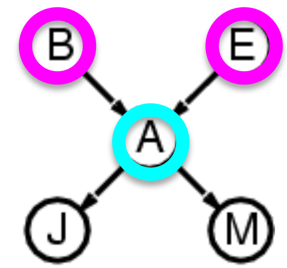
- John and Mary are conditionally independent of Burglary and Earthquake given Alarm
  - *Children* are conditionally independent of *ancestors* given *parents*

Conditional independence relationships



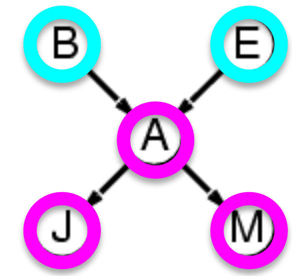
- John and Mary are conditionally independent of Burglary and Earthquake given Alarm
  - *Children* are conditionally independent of *ancestors* given *parents*
- John and Mary are conditionally independent of each other given Alarm
  - *Siblings* are conditionally independent of each other given *parents*

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 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*
- Alarm is *not* conditionally independent of John and Mary given Burglary and Earthquake
  - Nodes are *not* conditionally independent of *children* given *parents*

- **General rule:** each node is conditionally independent of its *non-descendants* given its *parents*

Conditional independence and the joint distribution

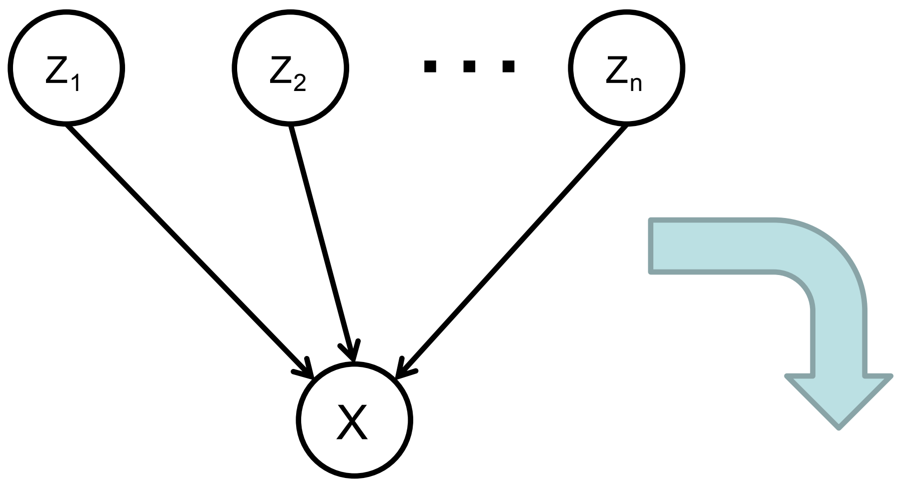
- **General rule:** each node is conditionally independent of its *non-descendants* given its *parents*
- Suppose the nodes $X_1, \ldots, X_n$ are sorted in topological order (parents before children)
- To get the joint distribution $P(X_1, \ldots, X_n)$, use chain rule:

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

$$= \prod_{i=1}^{n} P\left(X_i \mid Parents(X_i)\right)$$

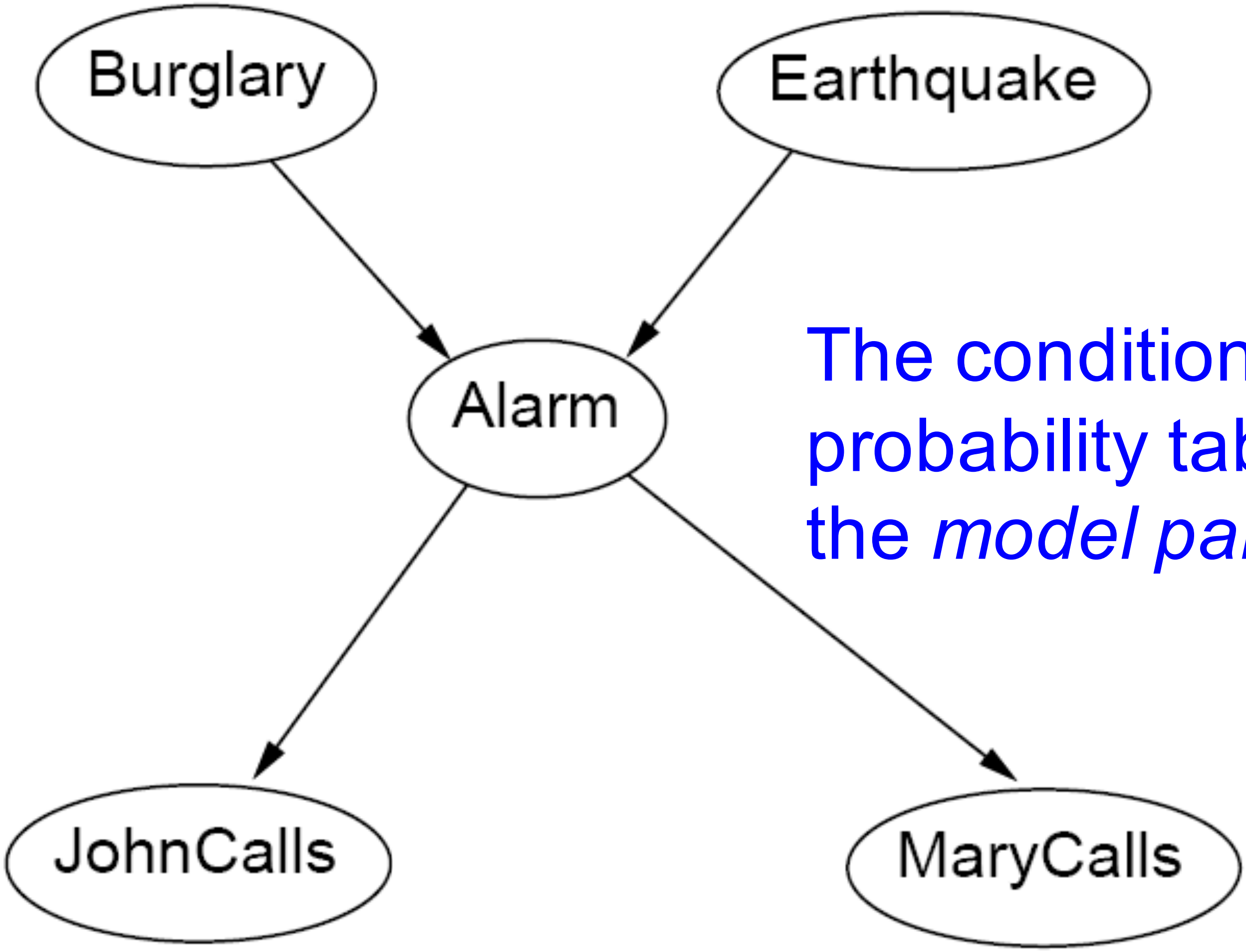
Conditional probability distributions

- To specify the full joint distribution, we need to specify a *conditional* distribution for each node given its parents: P (X | Parents(X))



$P(X \mid Z_1, \ldots, Z_n)$

Example: Burglar Alarm

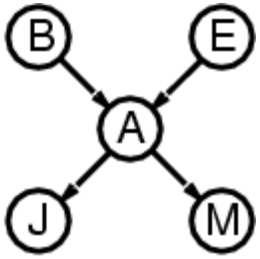


The conditional probability tables are the *model parameters*

The joint probability distribution

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

- For example, P(j, m, a, ¬b, ¬e)
  = P(¬b) P(¬e) P(a | ¬b, ¬e) P(j | a) P(m | a)



Conditional independence

- General rule: X is conditionally independent of every *non-descendant node* given its *parents*
- Example: *causal chain*

X: Low pressure

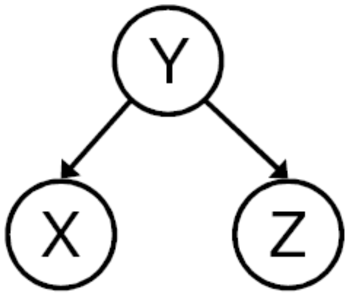
Y: Rain

Z: Traffic

- Are X and Z independent?
- Is Z independent of X given Y?

Conditional independence

- Common cause

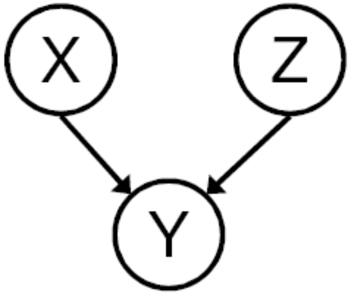


Y: Project due

X: Newsgroup busy

Z: Lab full

- Common effect



X: Raining

Z: Ballgame

Y: Traffic

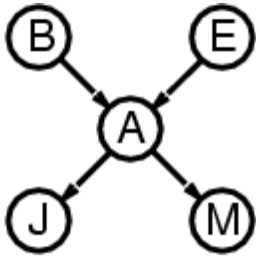
- Are X and Z independent?
  - No
- Are they conditionally independent given Y?
  - Yes

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

Compactness

- Suppose we have a Boolean variable X_i with k Boolean parents. How many rows does its conditional probability table have?
  - 2^k rows for all the combinations of parent values
  - Each row requires one number for $P(X_i = true \mid parent\ values)$
- If each variable has no more than k parents, how many numbers does the complete network require?
  - $O(n \cdot 2^k)$ numbers – vs. $O(2^n)$ for the full joint distribution
- How many numbers for the burglary network?

  1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)



Constructing Bayesian networks

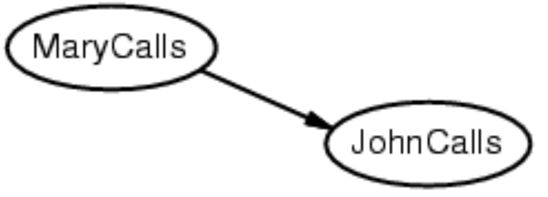
1. Choose an ordering of variables $X_1, \dots , X_n$
2. For i = 1 to n
   - add X_i to the network
   - select parents from $X_1, \dots ,X_{i-1}$ such that
     $P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \dots X_{i-1})$

Example

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

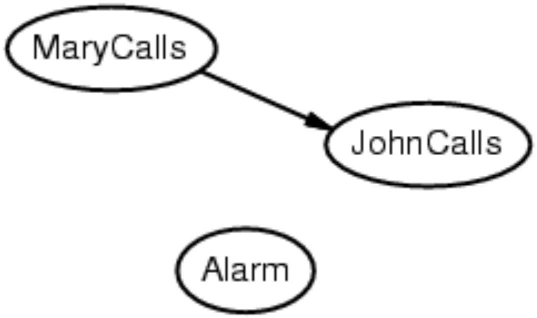
Example

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



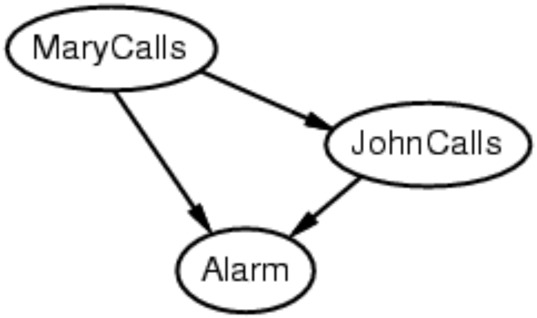
Example

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



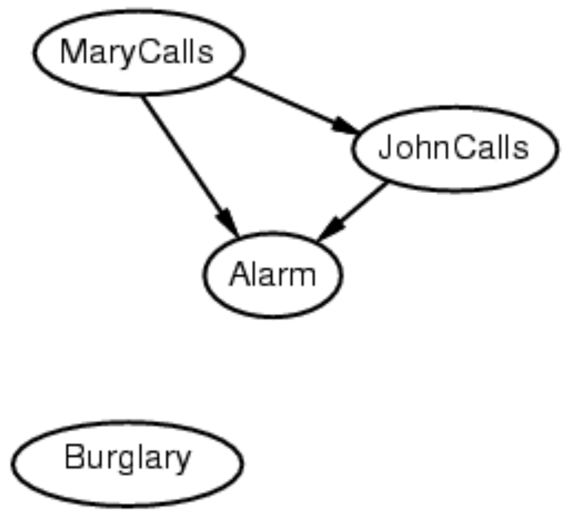
Example

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



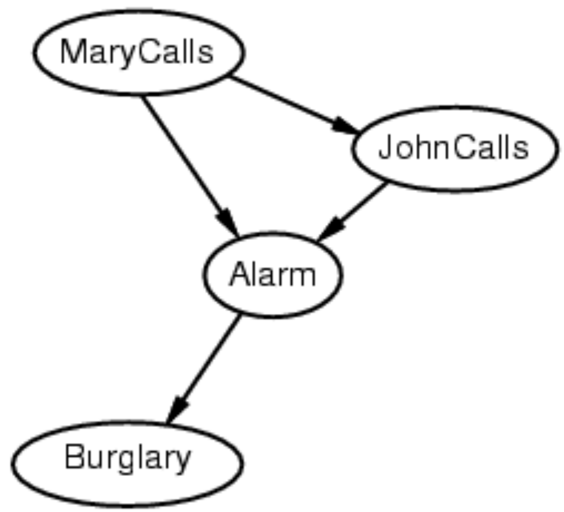
Example

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



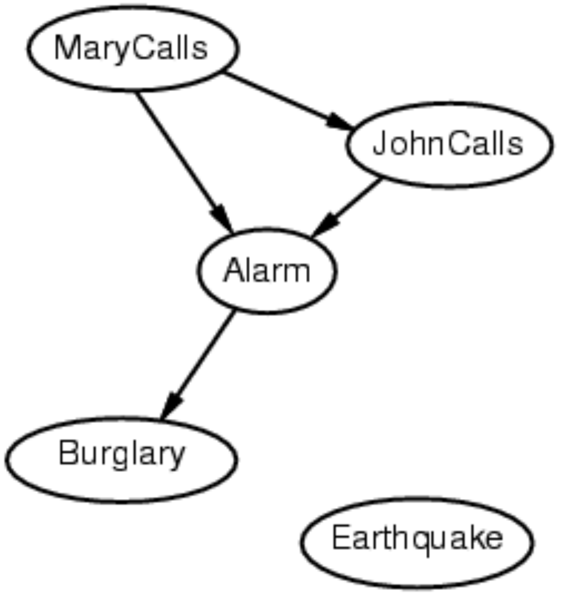
Example

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



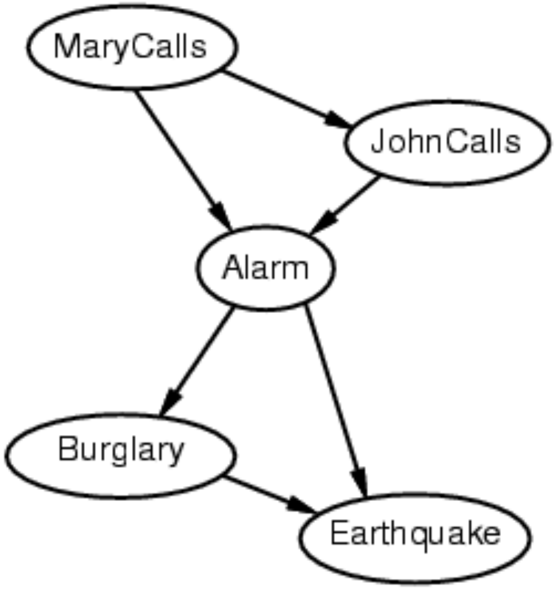
Example

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

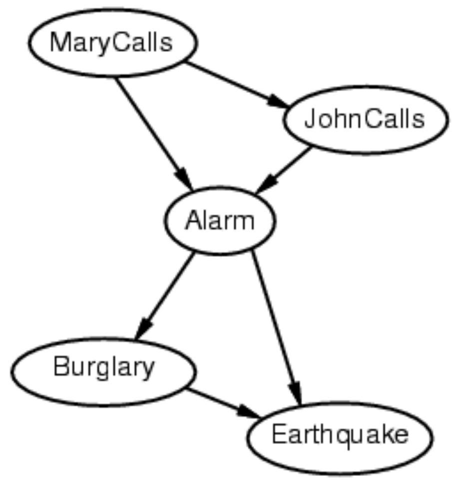


Example

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



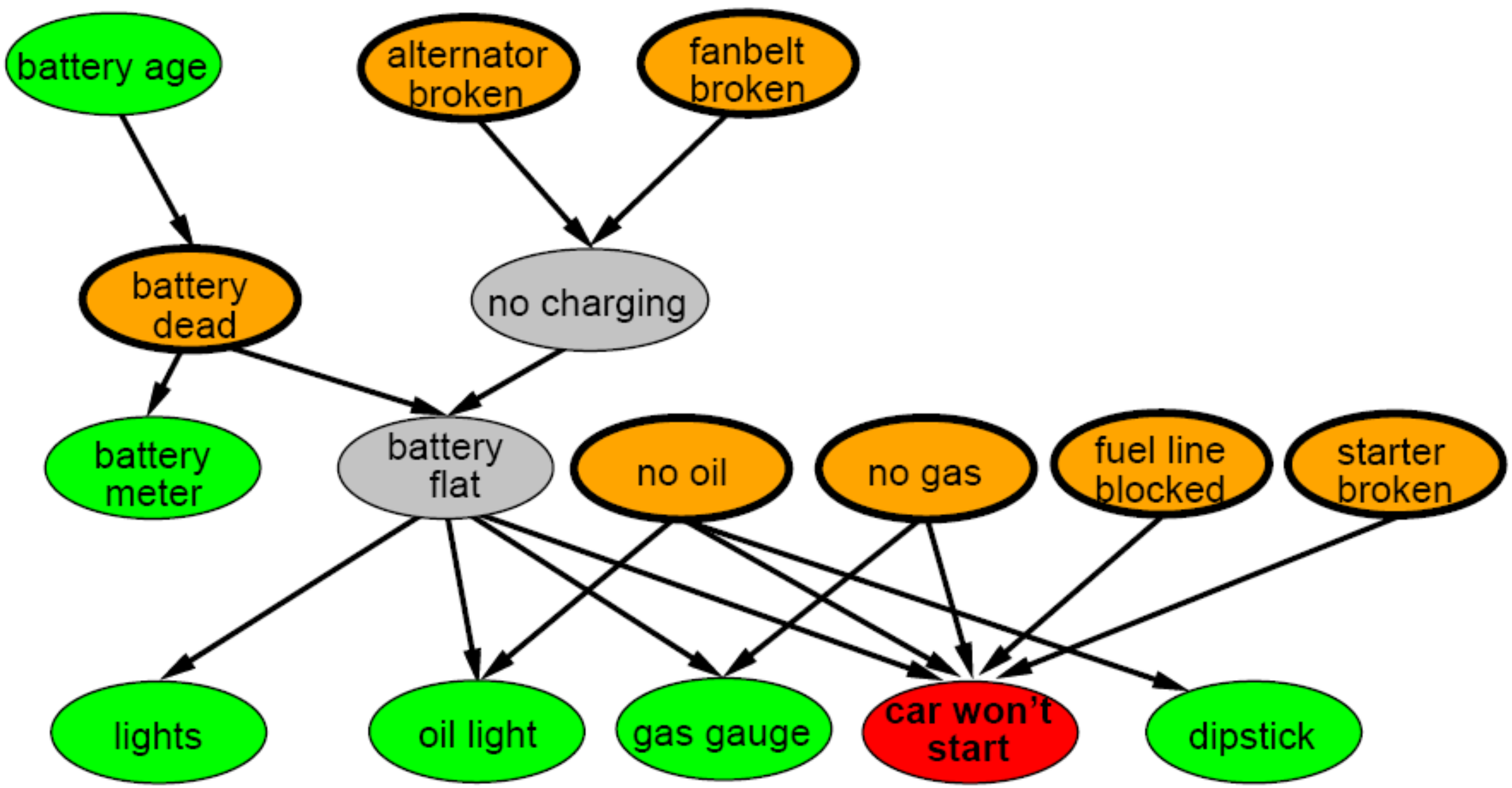
Example contd.



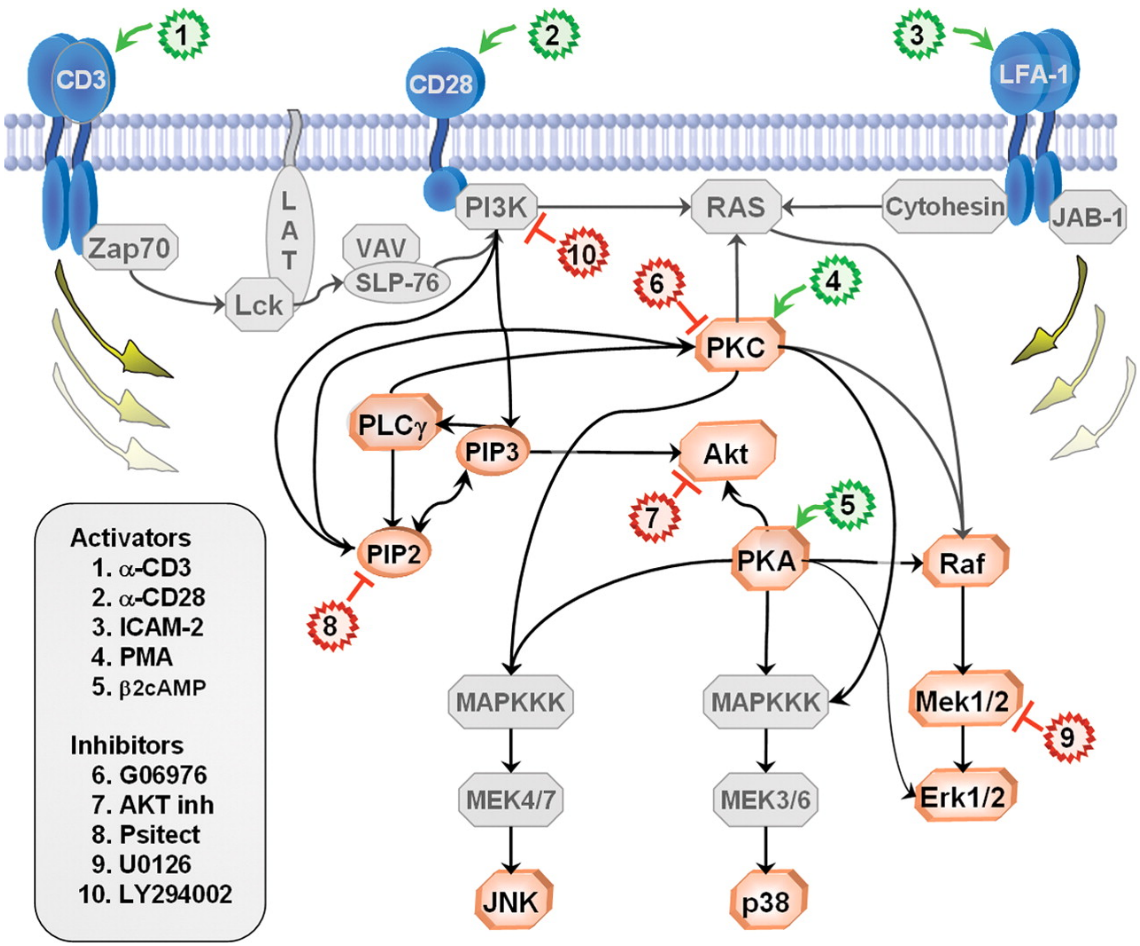
- Deciding conditional independence is hard in noncausal directions
  - The causal direction seems much more natural
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed (vs. 10 for the causal ordering)

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



In research literature…



Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data
Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan
(22 April 2005) *Science* **308** (5721), 523.

In research literature…

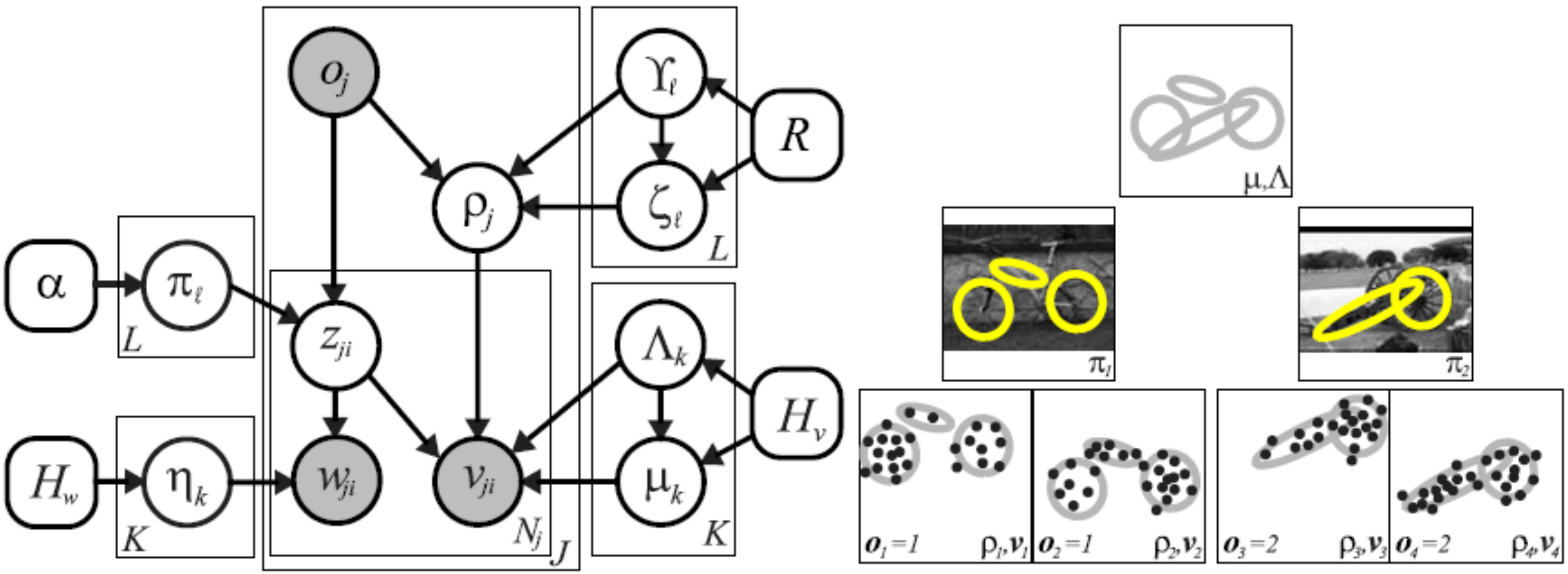


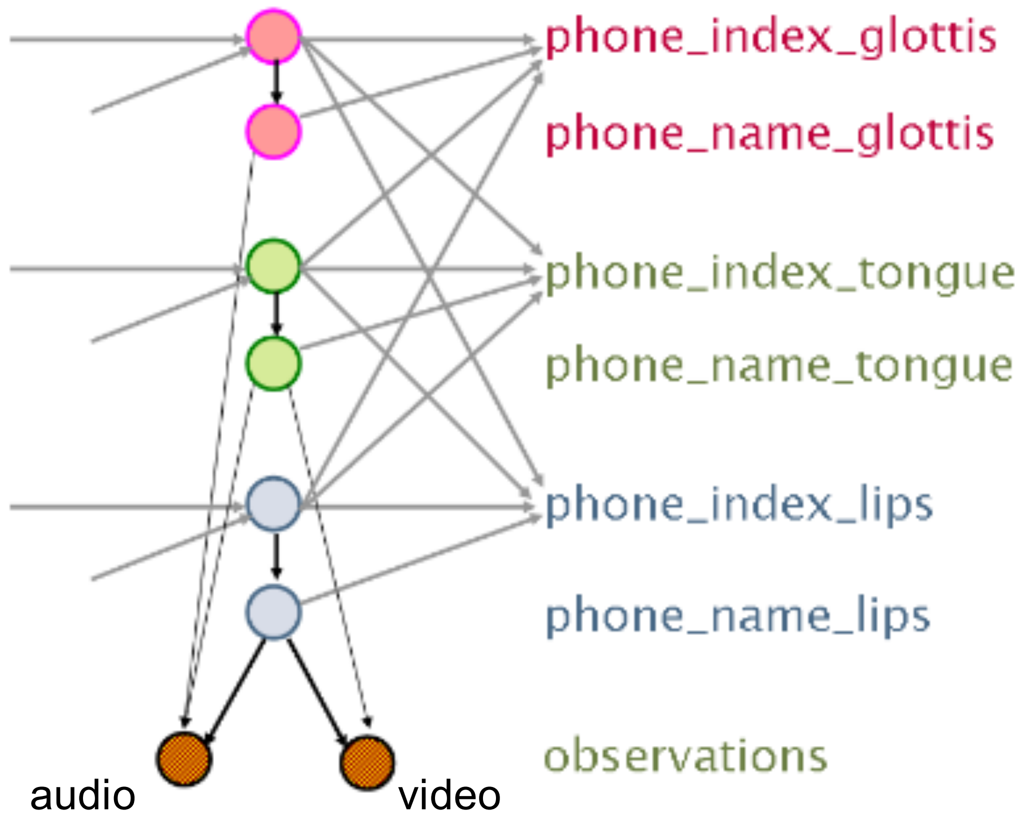
Fig. 3 A parametric, fixed-order model which describes the visual appearance of L object categories via a common set of K shared parts. The j^{th} image depicts an instance of object category o_j, whose position is determined by the reference transformation ρ_j. The appearance w_{ji} and position v_{ji}, relative to ρ_j, of visual features are determined by assignments $z_{ji} \sim \pi_{o_j}$ to latent parts. The cartoon example illustrates how a wheel part might be shared among two categories, *bicycle* and *cannon*. We show feature positions (but not appearance) for two hypothetical samples from each category

Describing Visual Scenes Using Transformed Objects and Parts

E. Sudderth, A. Torralba, W. T. Freeman, and A. Willsky.

International Journal of Computer Vision, No. 1-3, May 2008, pp. 291-330.

In research literature…



<u>**Audiovisual Speech Recognition with Articulator Positions as Hidden Variables**</u>

Mark Hasegawa-Johnson, Karen Livescu, Partha Lal and Kate Saenko

International Congress on Phonetic Sciences 1719:299-302, 2007

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + conditional probability tables
- Generally easy for domain experts to construct