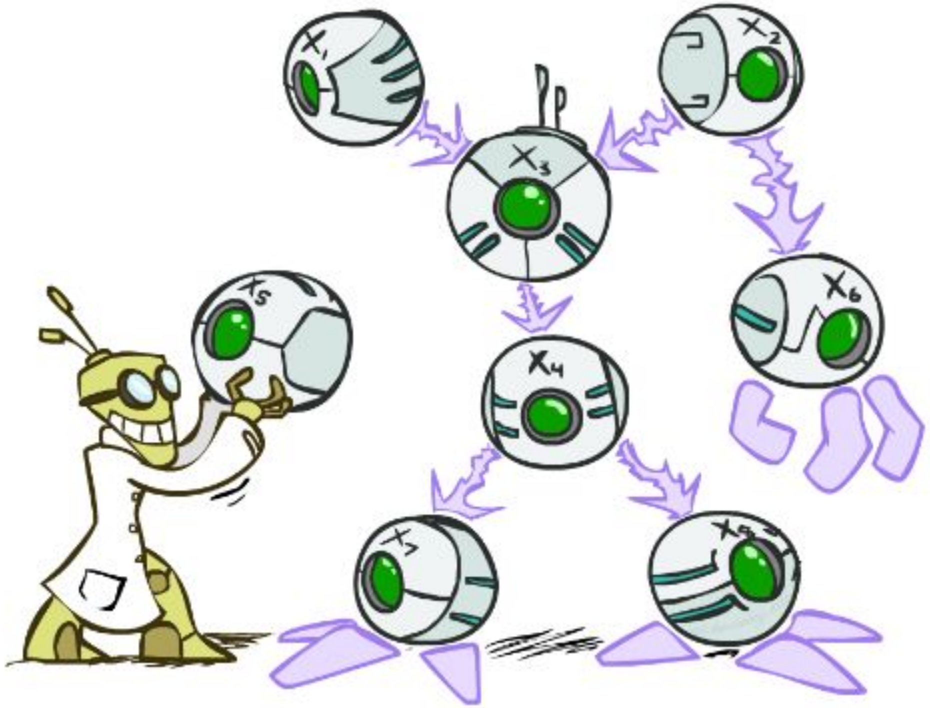
CS 188: Artificial Intelligence

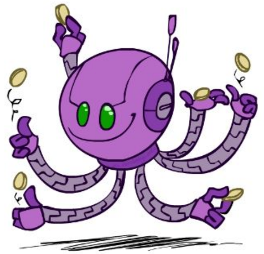
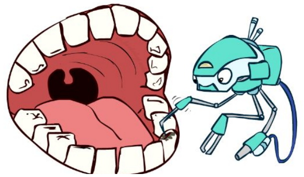
Bayes' Nets

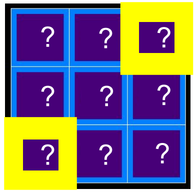


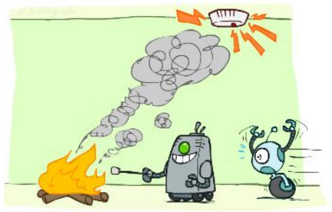
Instructors: Sergey Levine --- University of California, Berkeley

Reminders

- A probability model specifies a probability for every possible world
  - Typically, possible worlds are defined by assignments to a set of variables $X_1, \dots, X_n$
  - In that case, the probability model is a joint distribution $P(X_1, \dots, X_n)$
  - Written as a table, this would be exponential in n
- Independence: joint distribution = product of marginal distributions
  - $P(x, y) = P(x)P(y)$ or $P(x) = P(x \mid y)$
  - E.g., probability model for n coins represented by n numbers instead of 2^n
- Independence is rare in practice: within a domain, most variables correlated
- Conditional independence is much more common:
  - Toothache and Catch are conditionally independent given Cavity
  - Traffic and Umbrella are conditionally independent given Rain
  - Alarm and Fire are conditionally independent given Smoke
  - Reading1 and Reading2 are conditionally independent given Ghost location

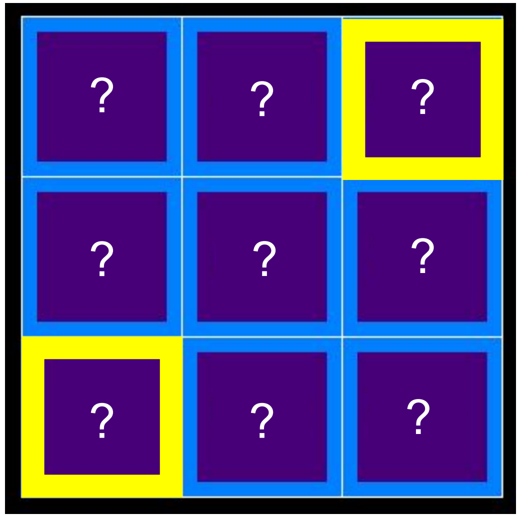


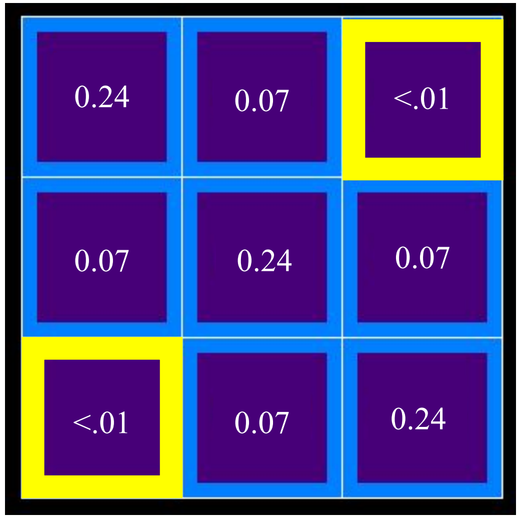


Ghostbusters, Revisited

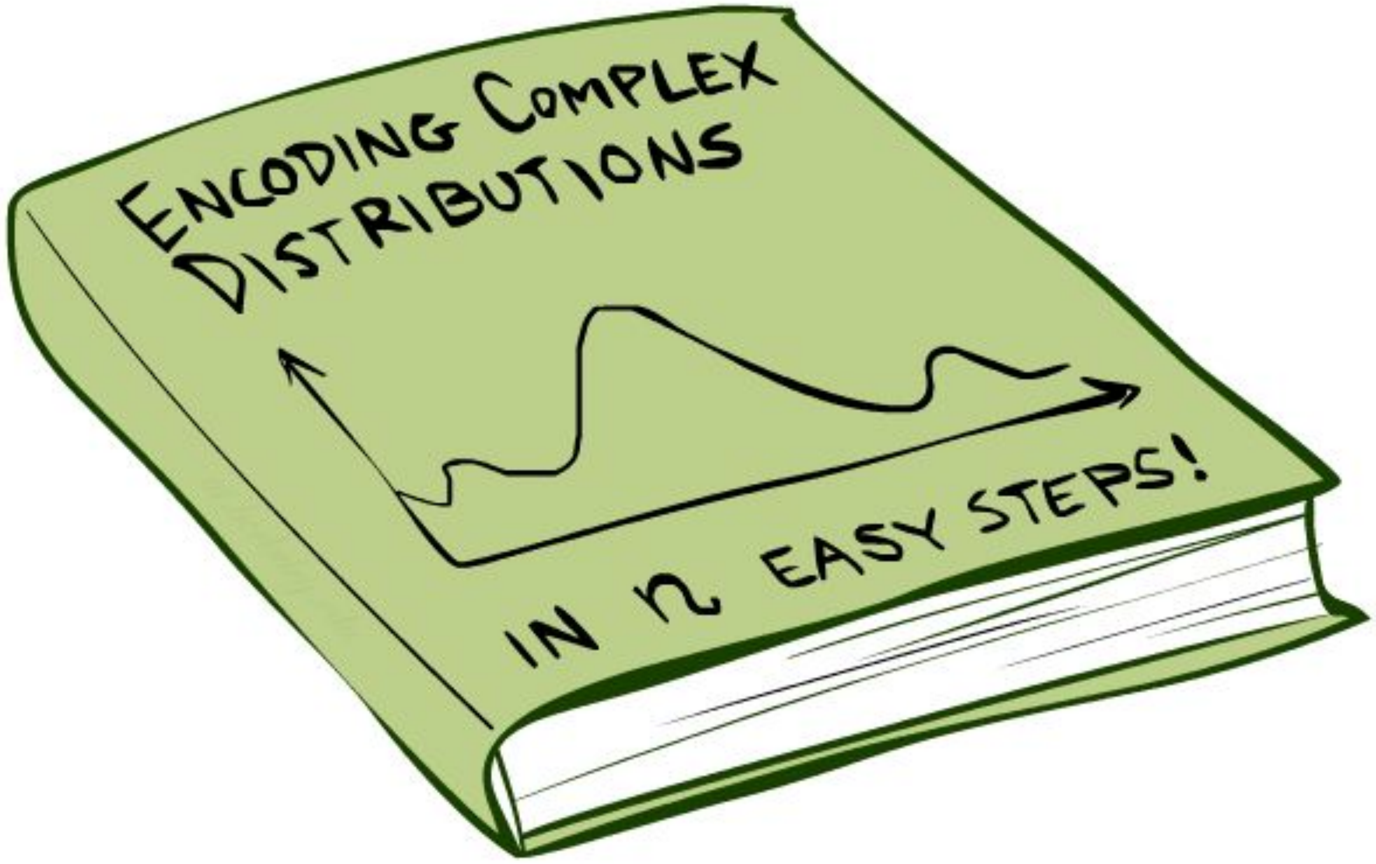
- What about two readings? What is $P(r_1,r_2 \mid g)$?
- Readings are conditionally independent given the ghost location!
- $P(r_1,r_2 \mid g) = P(r_1 \mid g)\, P(r_2 \mid g)$
- Applying Bayes' rule in full:
- $P(g \mid r_1,r_2)$ α $P(r_1,r_2 \mid g)\, P(g)$

  $= P(g)\, P(r_1 \mid g)\, P(r_2 \mid g)$
- _Bayesian updating_ using low-dimensional conditional distributions!!



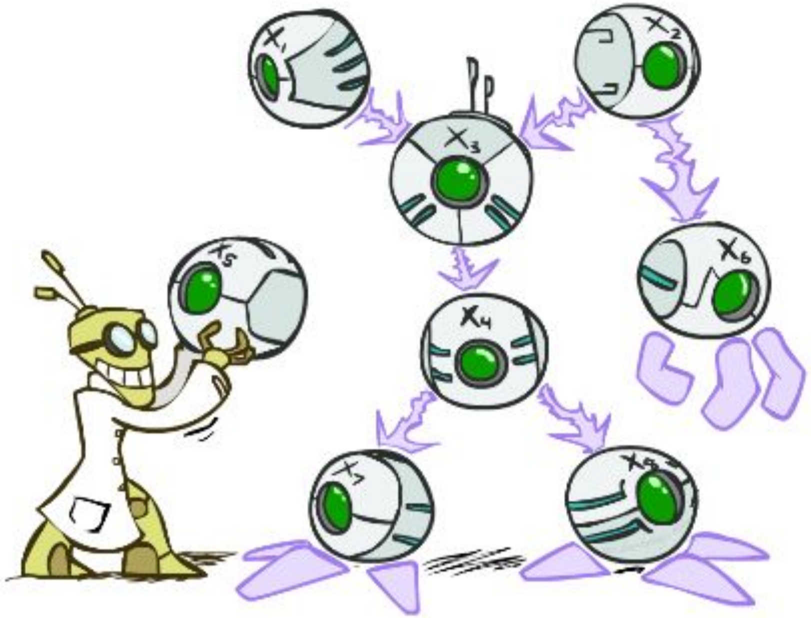


Bayes Nets: Big Picture



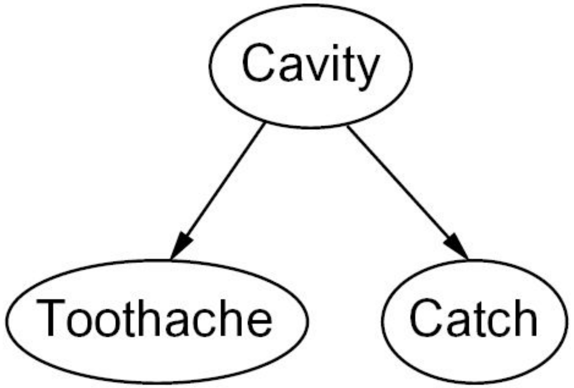
Bayes Nets: Big Picture

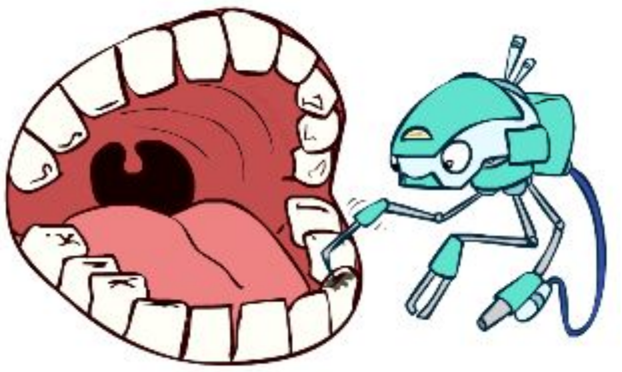
- Bayes nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - A subset of the general class of graphical models
- Take advantage of local causality:
  - the world is composed of many variables,
  - each interacting locally with a few others
- For about 10 min, we'll be vague about how these interactions are specified



Graphical Model Notation

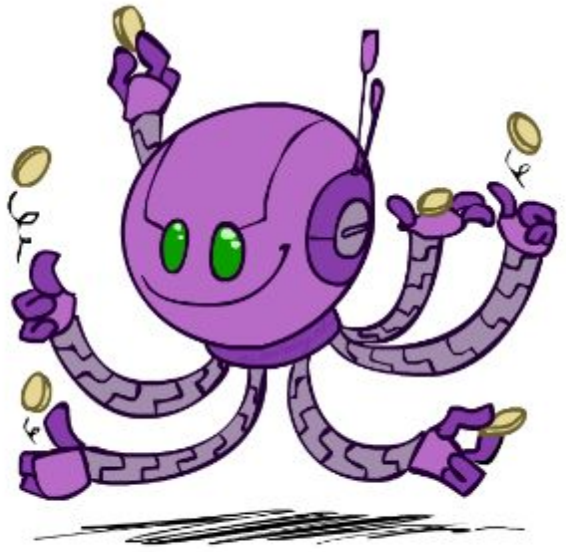
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don't!)





Example: Coin Flips

- N independent coin flips

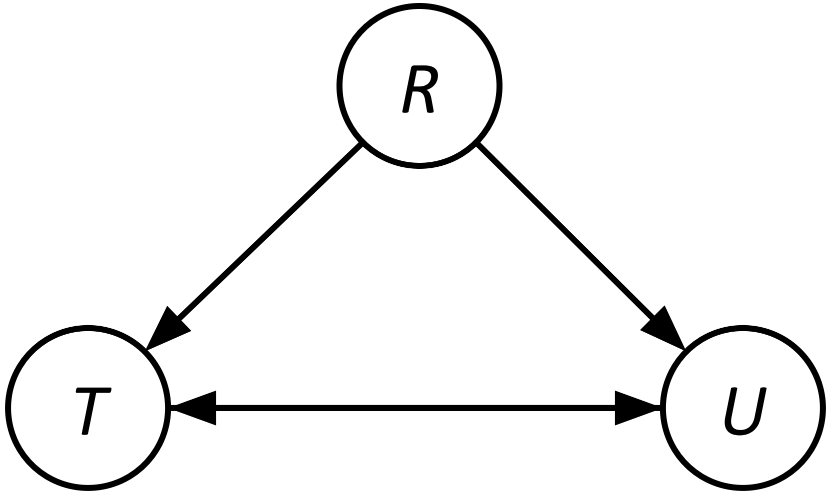


X_1 X_2 $\cdots$ X_n

- No interactions between variables: absolute independence

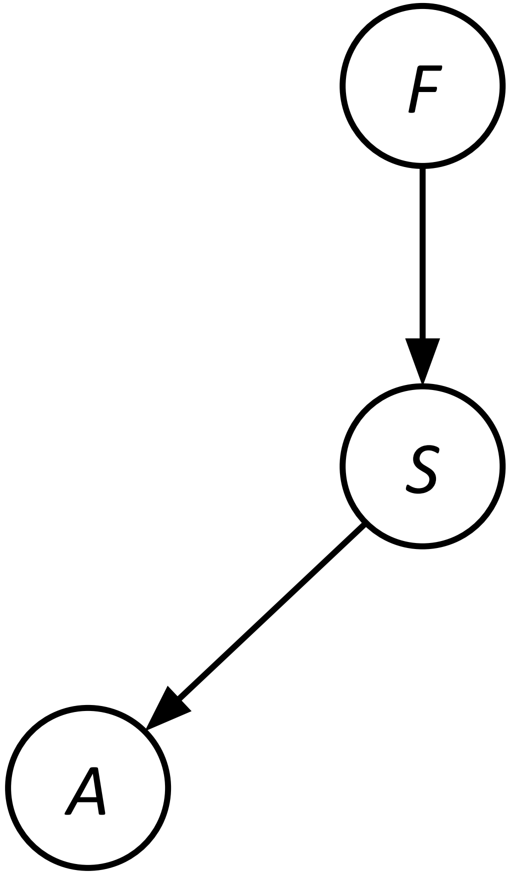
Example: Traffic

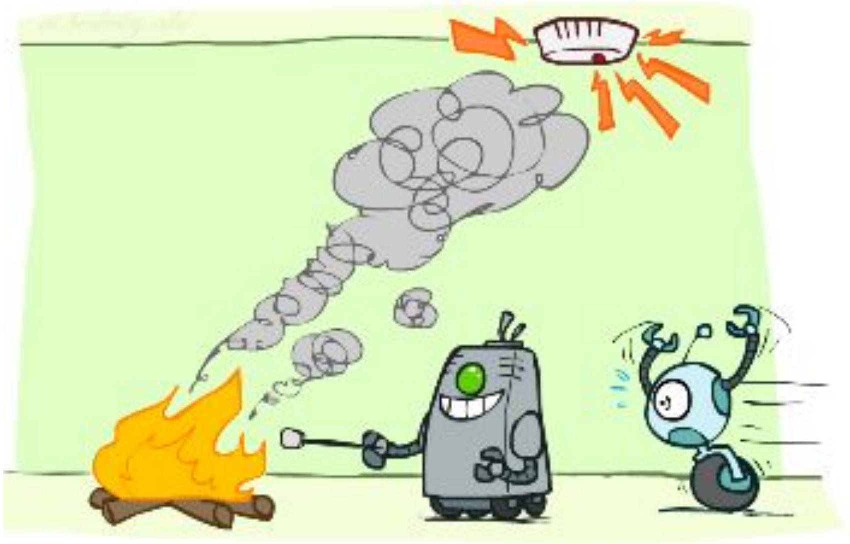
- Variables:
  - T: There is traffic
  - U: I'm holding my umbrella
  - R: It rains



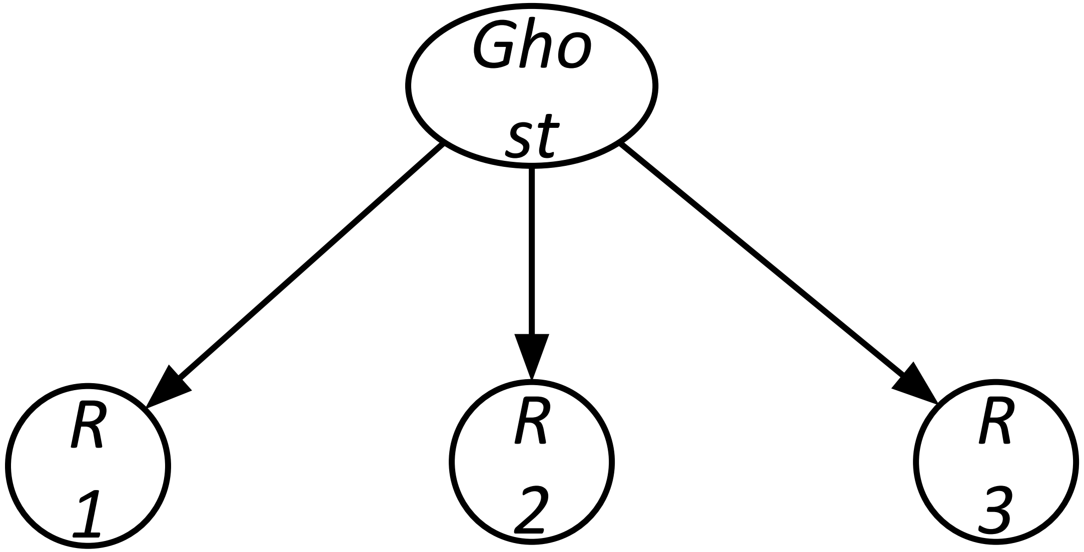
Example: Smoke alarm

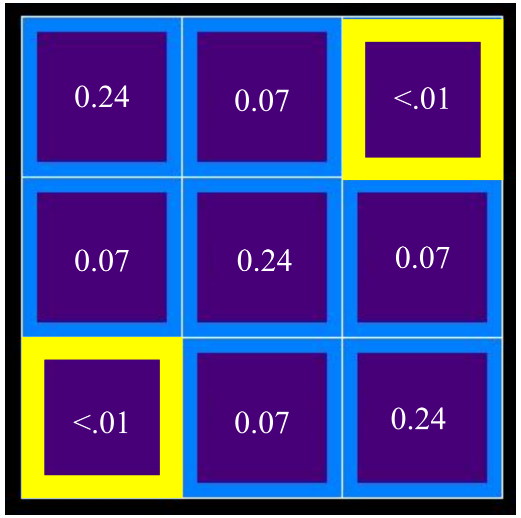
- Variables:
  - F: There is fire
  - S: There is smoke
  - A: Alarm sounds



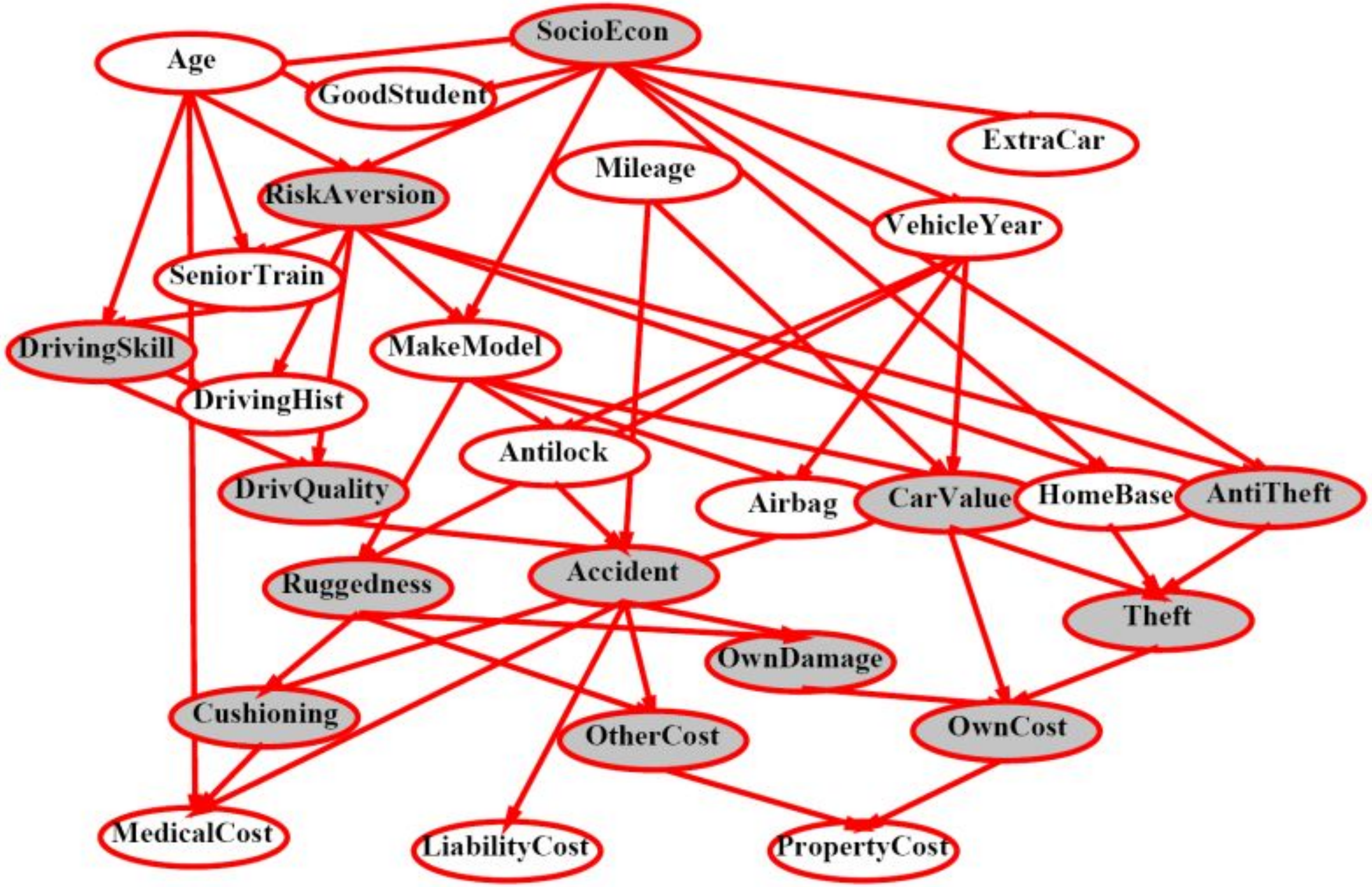


Example: Ghostbusters

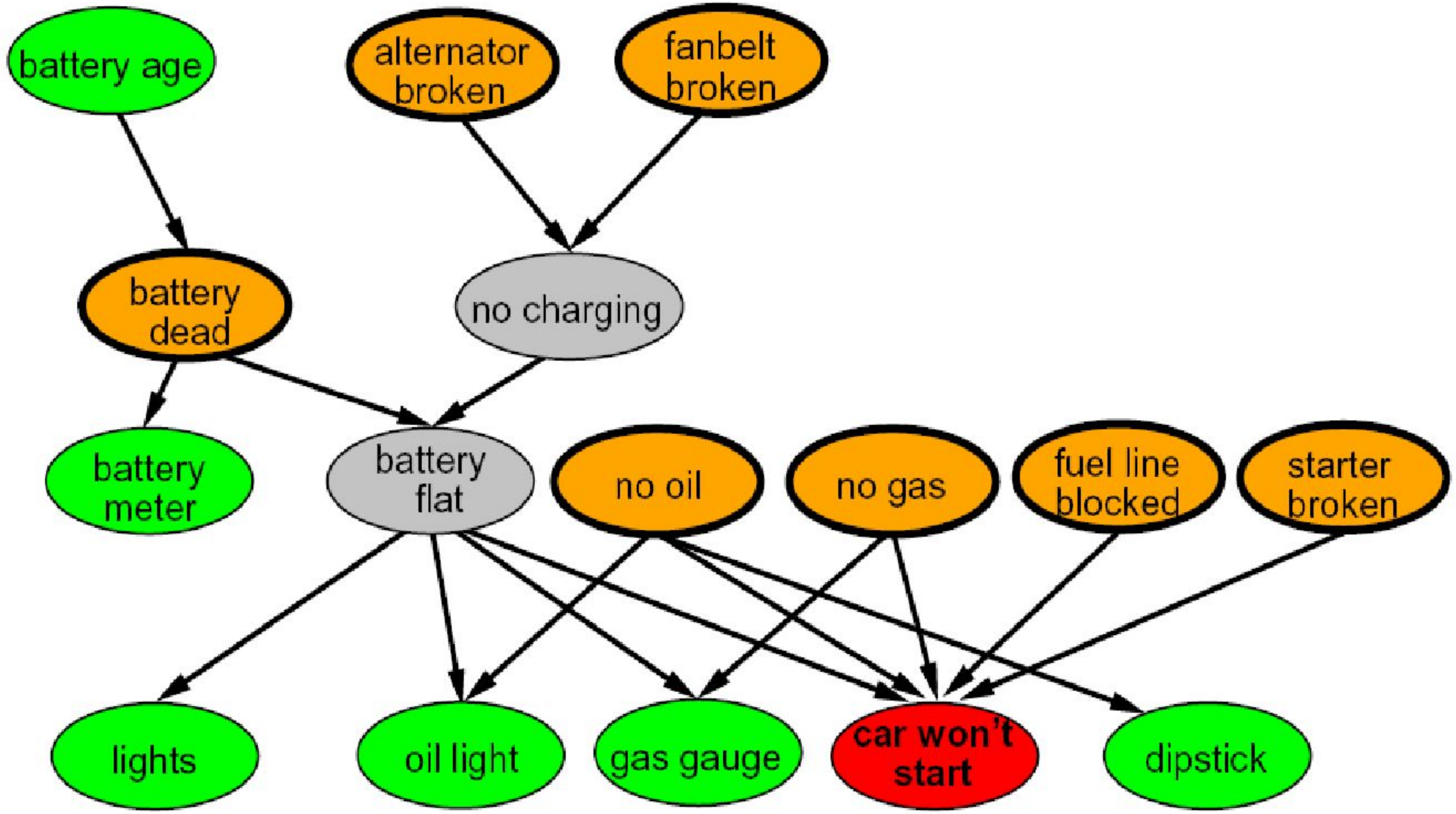




Example Bayes' Net: Insurance



Example Bayes' Net: Car

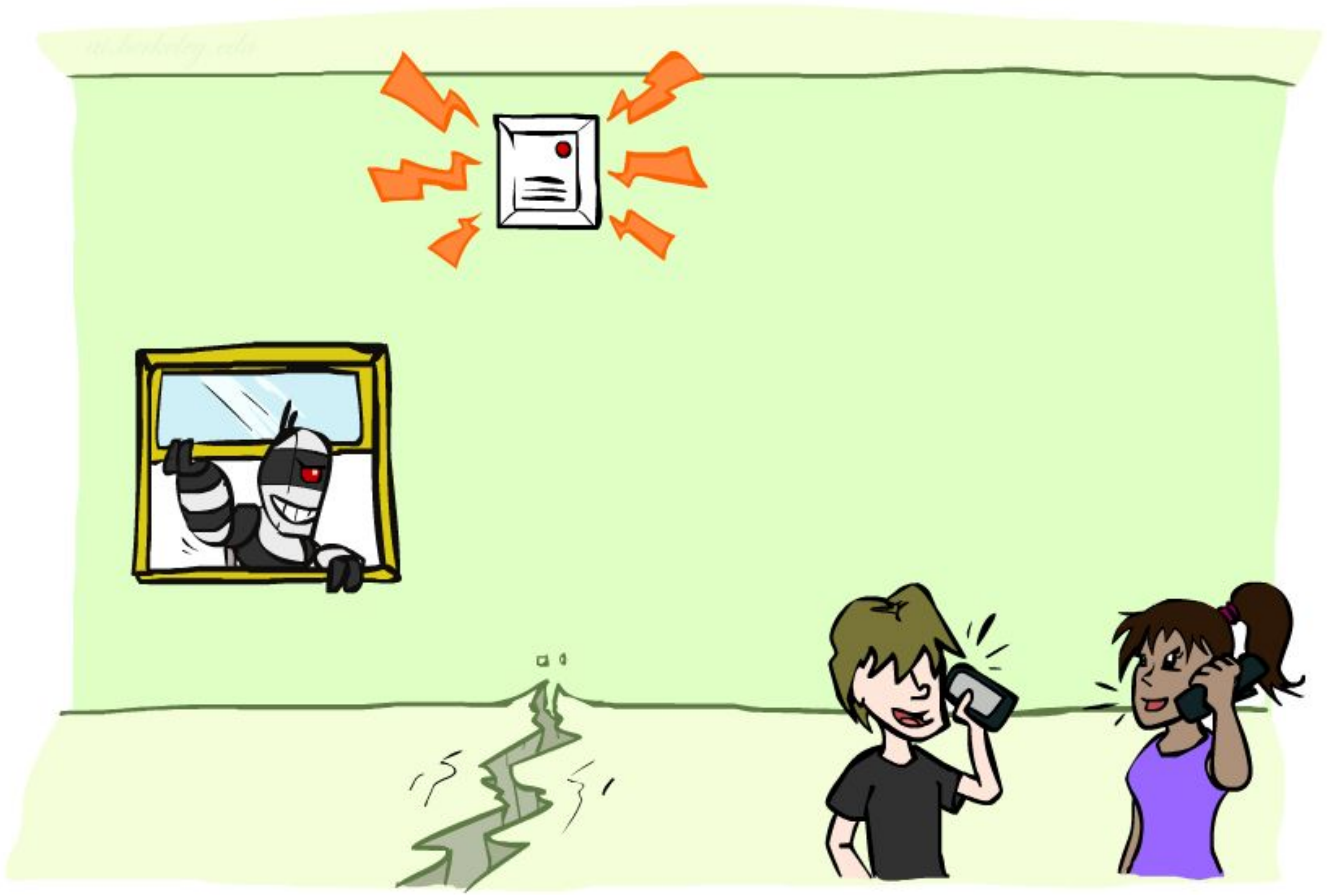


Can we build it?

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Can we build it?

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



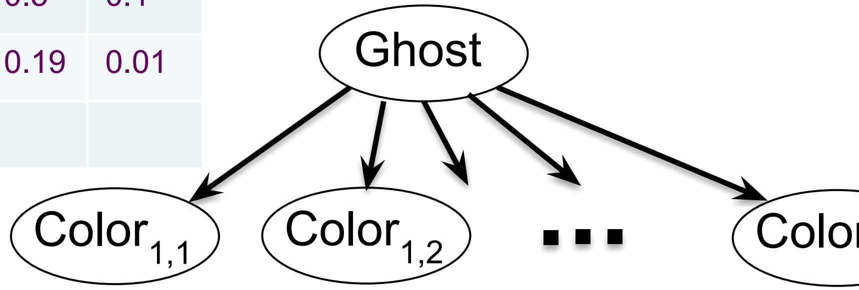
Bayes Net Syntax and Semantics

Bayes Net Syntax

- A set of nodes, one per variable X_i
- A directed, acyclic graph
- A conditional distribution for each node given its ***parent variables*** in the graph
  - ***CPT***: conditional probability table: each row is a distribution for child given a configuration of its parents
  - Description of a noisy "causal" process

Ghost	$P(Color_{1,1} \mid Ghost)$			
	g	y	o	r
(1,1)	0.01	0.1	0.3	0.59
(1,2)	0.1	0.3	0.5	0.1
(1,3)	0.3	0.5	0.19	0.01
…				

P(Ghost)			
(1,1)	(1,2)	(1,3)	…
0.11	0.11	0.11	…

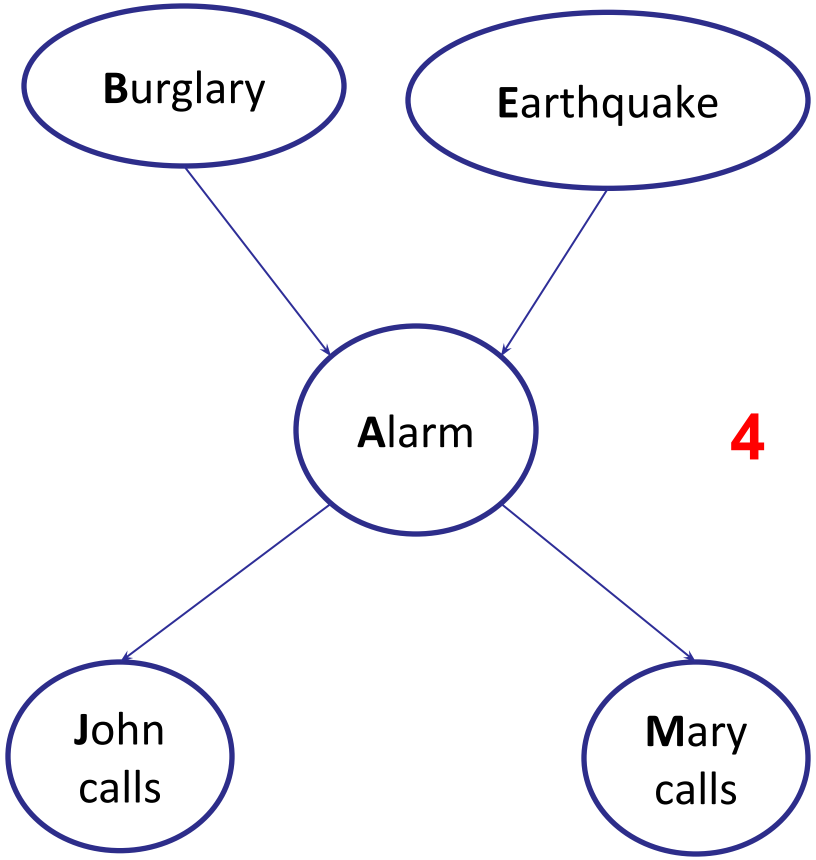


A Bayes net = Topology (graph) + Local Conditional Probabilities

Example: Alarm Network

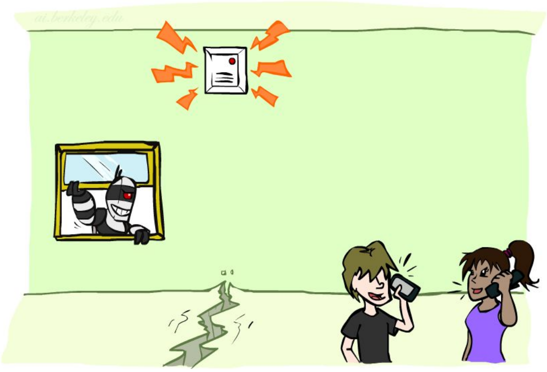
P(B)	
true	false
0.001	0.999

1



1

P(E)	
true	false
0.002	0.998



4

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

2

2

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Number of free parameters in each CPT:

Parent domain sizes $d_1,\ldots,d_k$

Child domain size d
Each table row must sum to 1

$(d-1)\ \Pi_i\ d_i$

General formula for sparse BNs

- Suppose
  - n variables
  - Maximum domain size is d
  - Maximum number of parents is k
- Full joint distribution has size $O(d^n)$
- Bayes net has size $O(n \cdot d^k)$
  - Linear scaling with n as long as causal structure is local

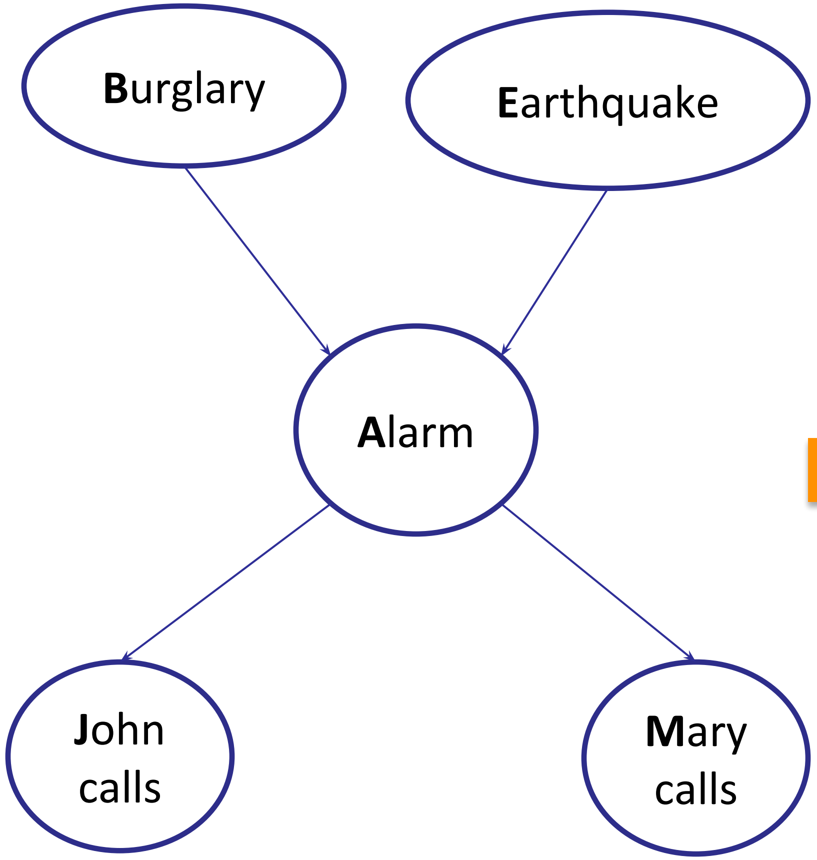
Bayes net global semantics

- Bayes nets encode joint distributions as product of conditional distributions on each variable:

$$P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$$

Example

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

P(b,¬e, a, ¬j, ¬m) =

P(b) P(¬e) P(a|b,¬e) P(¬j|a) P(¬m|a)

=.001x.998x.94x.1x.3=.000028

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Probabilities in BNs

- Why are we guaranteed that setting

$$P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$$

  results in a proper joint distribution?

- Chain rule (valid for all distributions): $P(X_1,..,X_n) = \prod_i P(X_i \mid X_1,\ldots,X_{i-1})$

- <u>Assume</u> conditional independences: $P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid Parents(X_i))$
  - When adding node X_i, ensure parents "shield" it from other predecessors

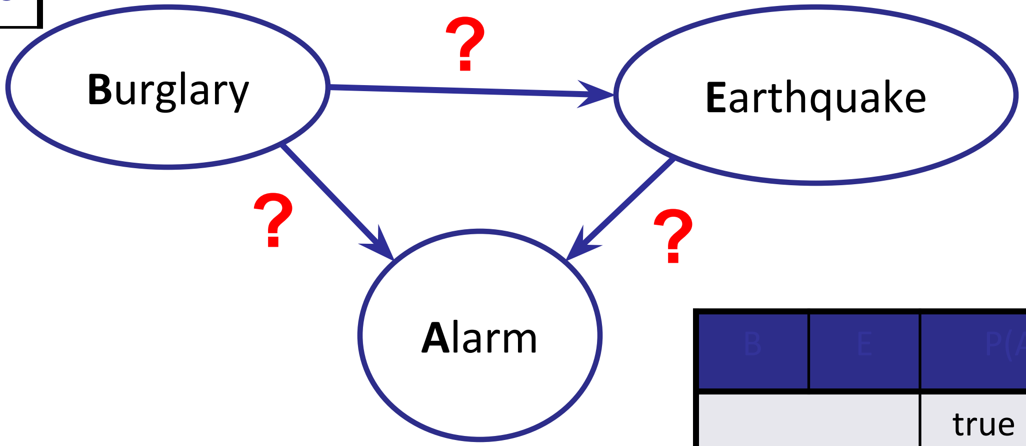
- Consequence: $P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$

- So the topology implies that certain conditional independencies hold

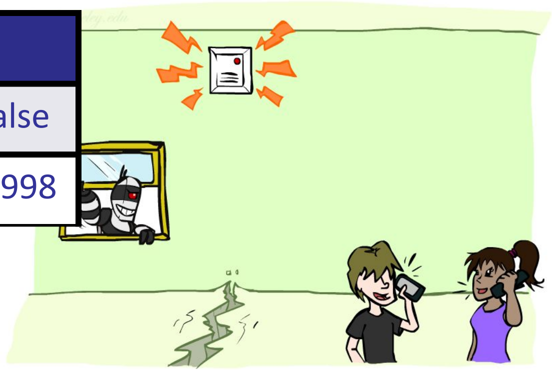
Example: Burglary

- Burglary
- Earthquake
- Alarm

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

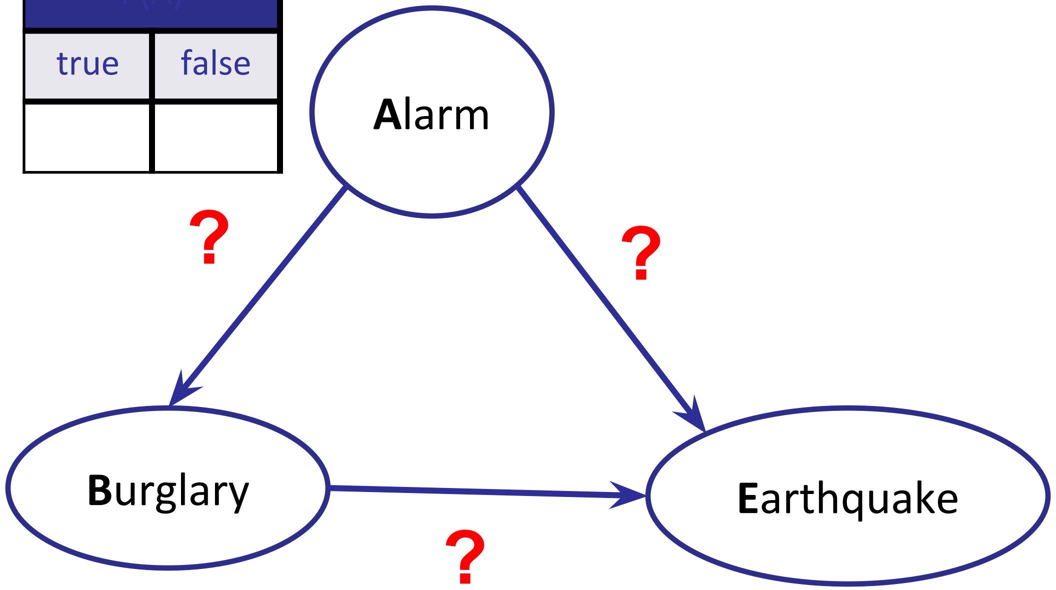


B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

Example: Burglary

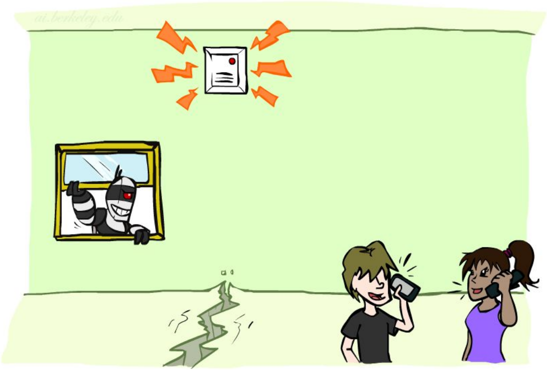
- Alarm
- Burglary
- Earthquake

P(A)	
true	false



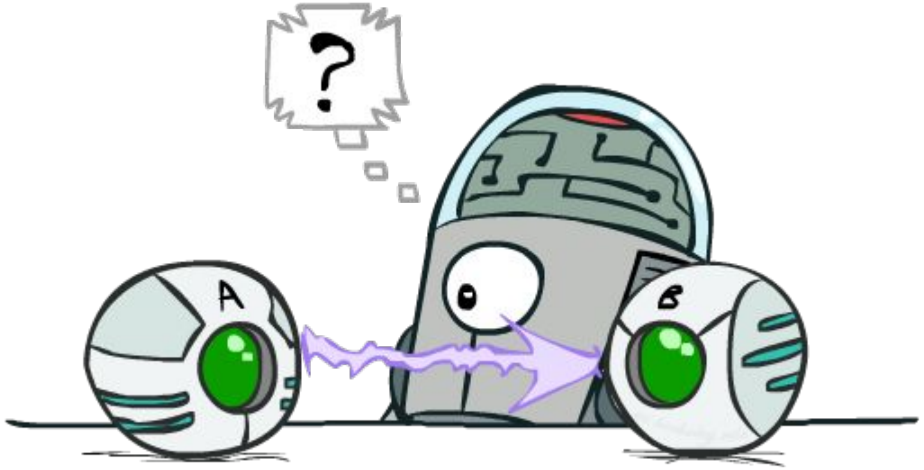
A	P(B|A)	
	true	false
true	?	
false		

A	B	P(E|A,B)	
		true	false
true	true	?	
true	false		
false	true		
false	false		



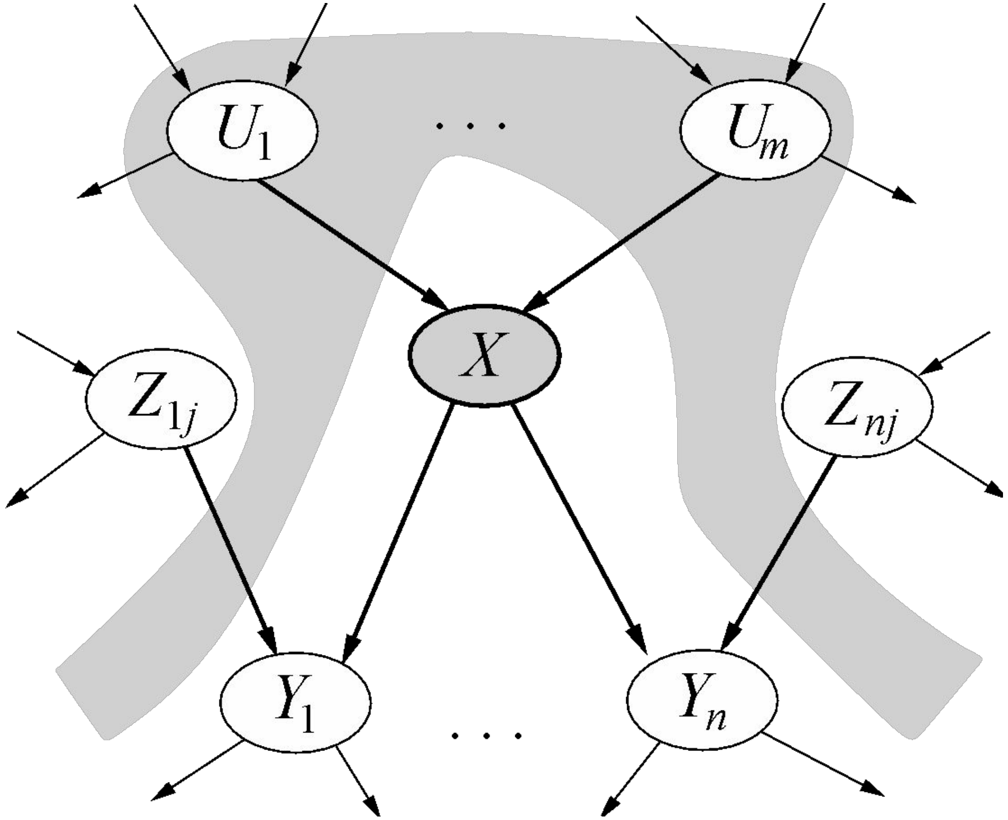
Causality?

- When Bayes nets reflect the true causal patterns:
  - Often simpler (fewer parents, fewer parameters)
  - Often easier to assess probabilities
  - Often more robust: e.g., changes in frequency of burglaries should not affect the rest of the model!
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *Umbrella*
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence:
    $P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid Parents(X_i))$



Conditional independence semantics

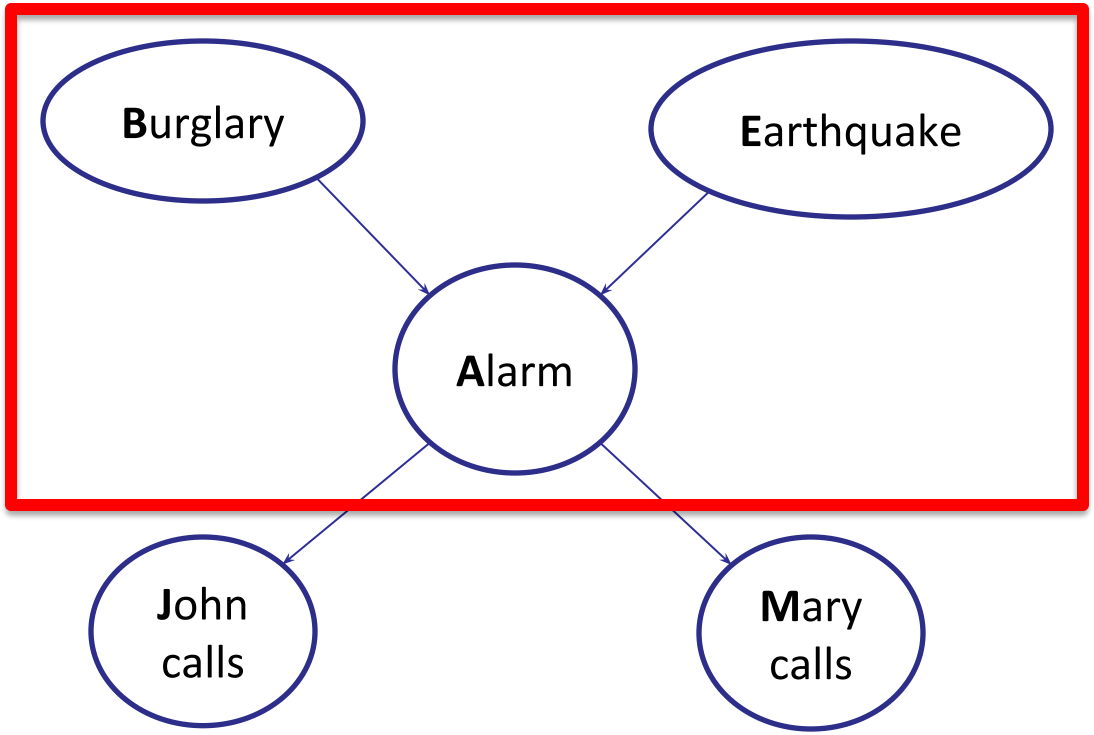
- ***Every variable is conditionally independent of its non-descendants given its parents***
- Conditional independence semantics <=> global semantics



Example

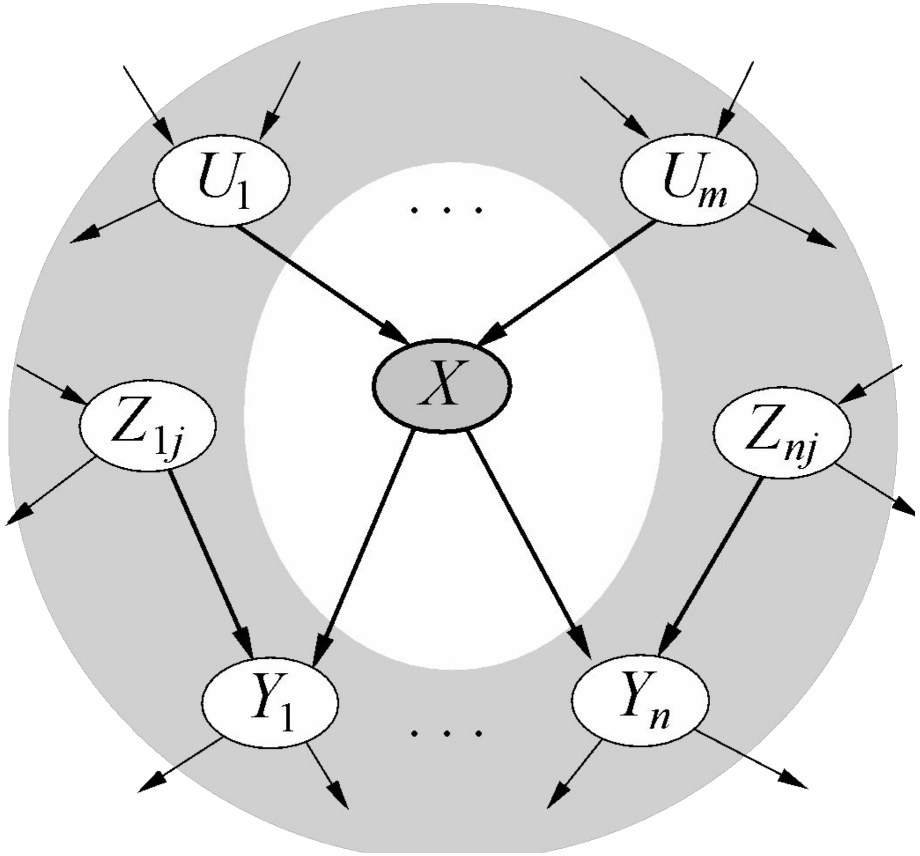
- JohnCalls independent of Burglary given Alarm?
  - Yes
- JohnCalls independent of MaryCalls given Alarm?
  - Yes
- Burglary independent of Earthquake?
  - Yes
- Burglary independent of Earthquake given Alarm?
  - NO!
  - Given that the alarm has sounded, both burglary and earthquake become more likely
  - But if we then learn that a burglary has happened, the alarm is ***explained away*** and the probability of earthquake drops back

V-structure



Markov blanket

- A variable's Markov blanket consists of parents, children, children's other parents
- ***Every variable is conditionally independent of all other variables given its Markov blanket***



Bayes Nets

- So far: how a Bayes net encodes a joint distribution
- Next: how to answer queries, i.e., compute conditional probabilities of queries given evidence