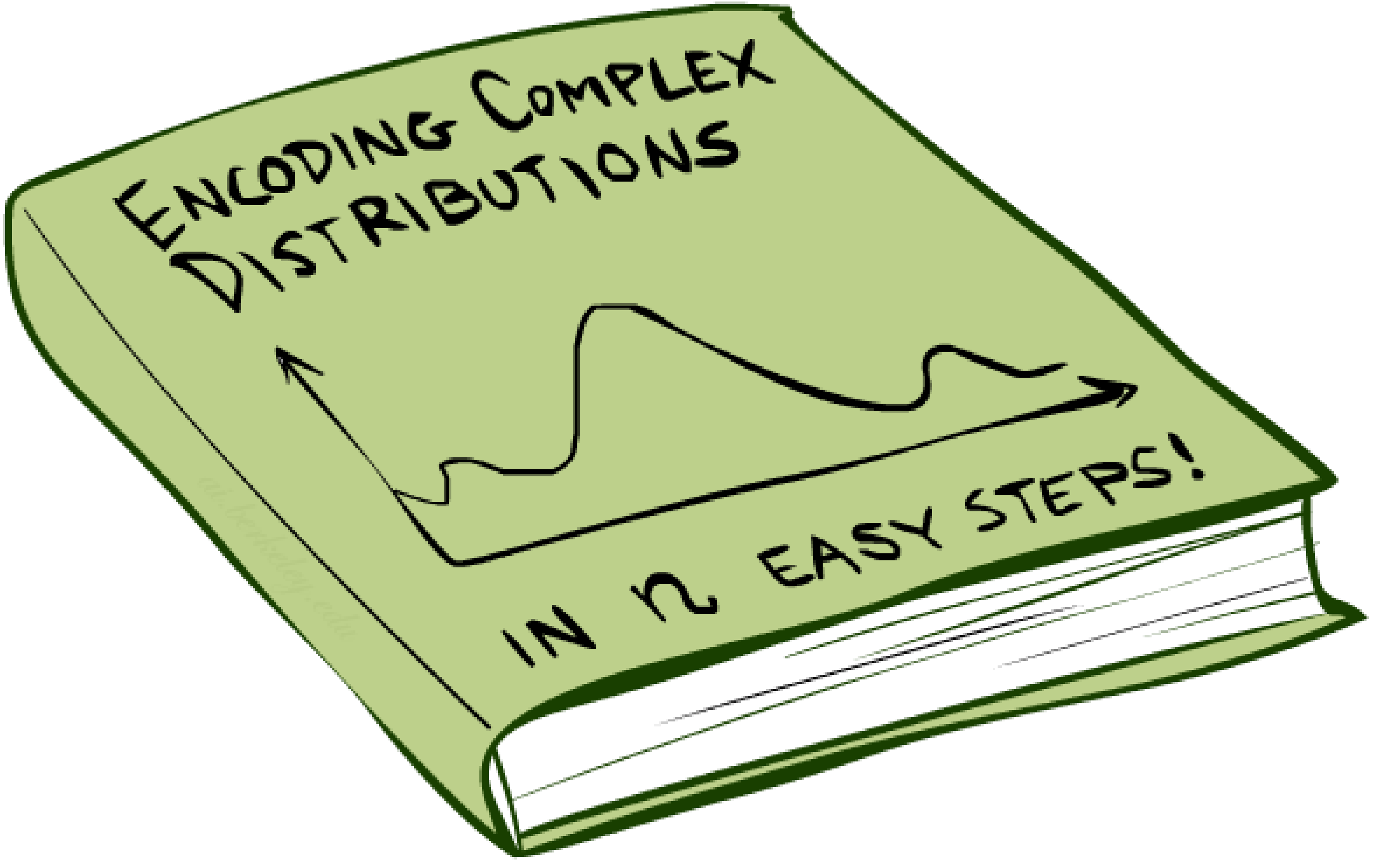
AI2002 – ARTIFICIAL INTELLIGENCE SPRING 2024 - LECTURE 34-36

Presented By: Mr. Sandesh Kumar

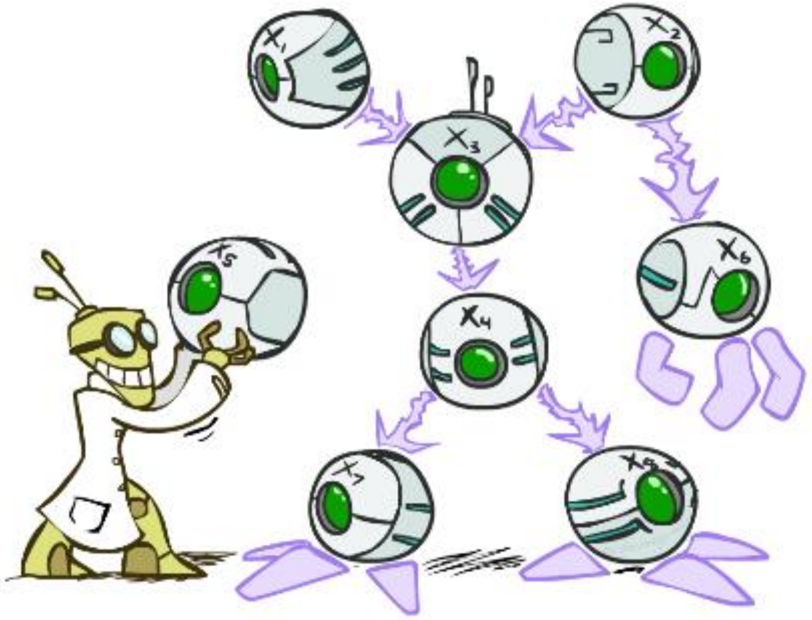
Slides are taken from UC Berkeley

Bayes Nets: Big Picture



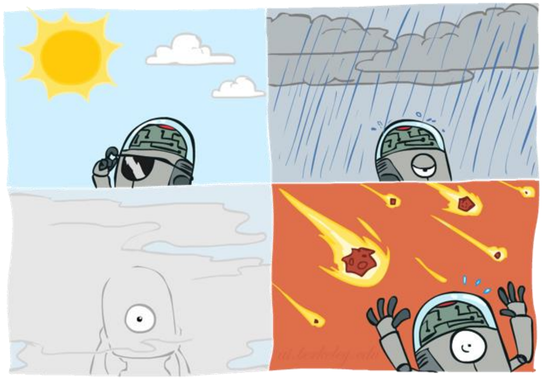
Bayes Nets: Big Picture

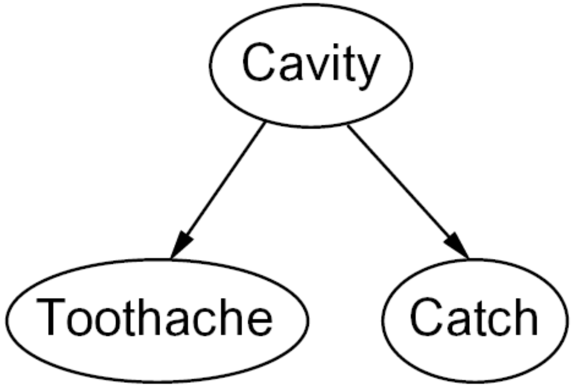
- Bayes nets: a technique for describing complex joint distributions (models) using simple, conditional distributions
  - A subset of the general class of graphical models
- Use local causality/conditional independence:
  - the world is composed of many variables,
  - each interacting locally with a few others
- Outline
  - Representation
  - Exact inference
  - Approximate inference

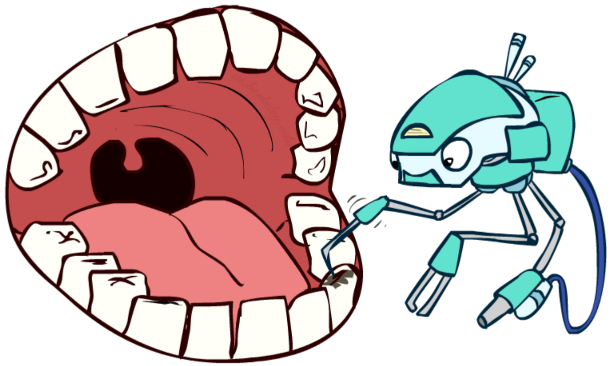


Graphical Model Notation

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

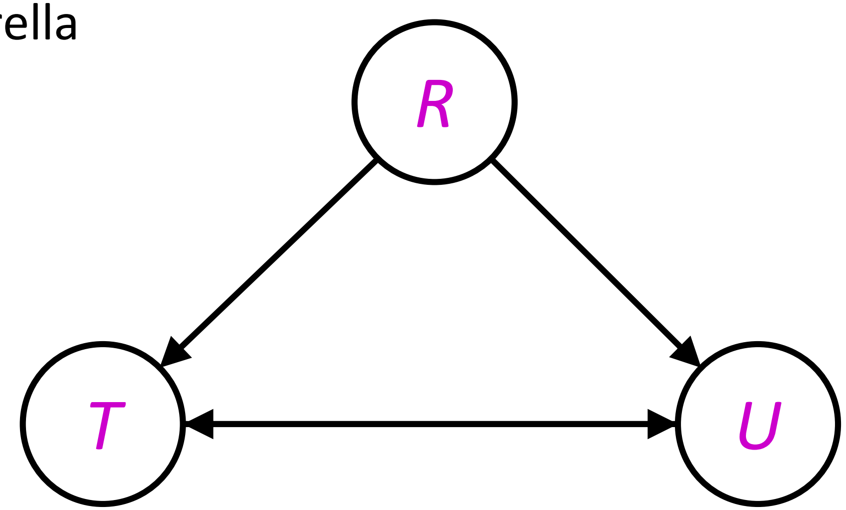


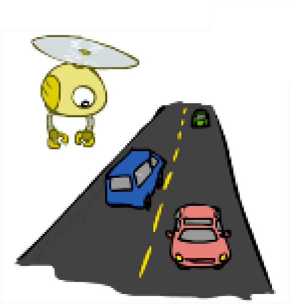


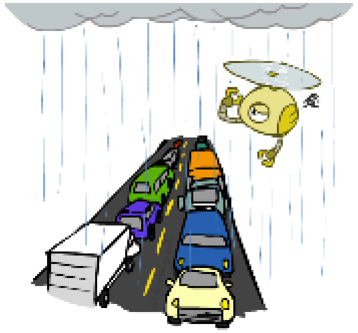


Example: Traffic

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

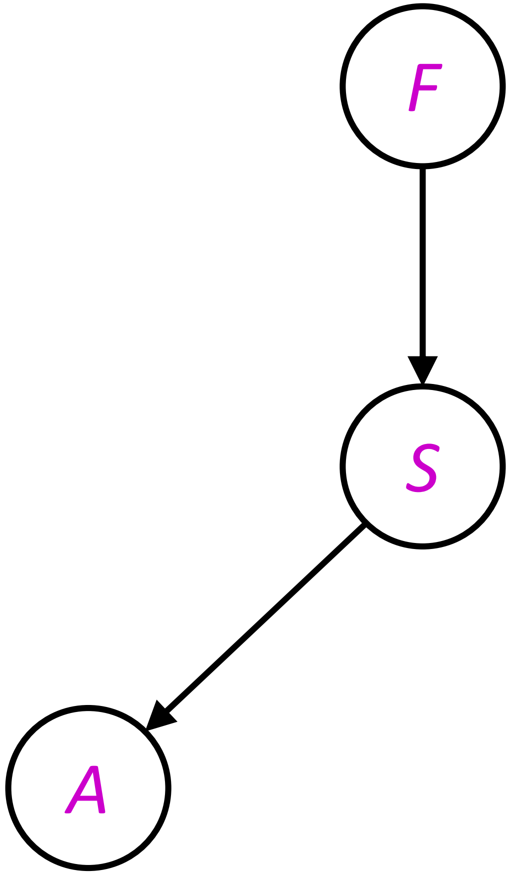


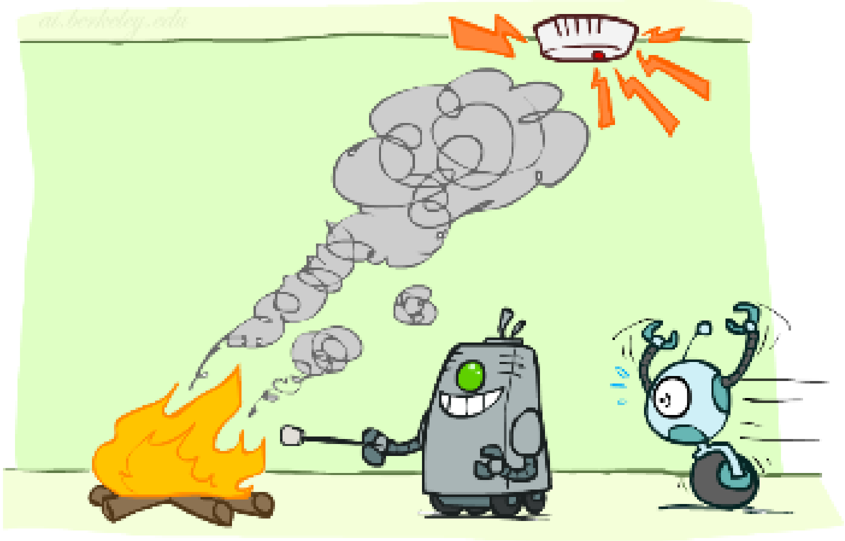


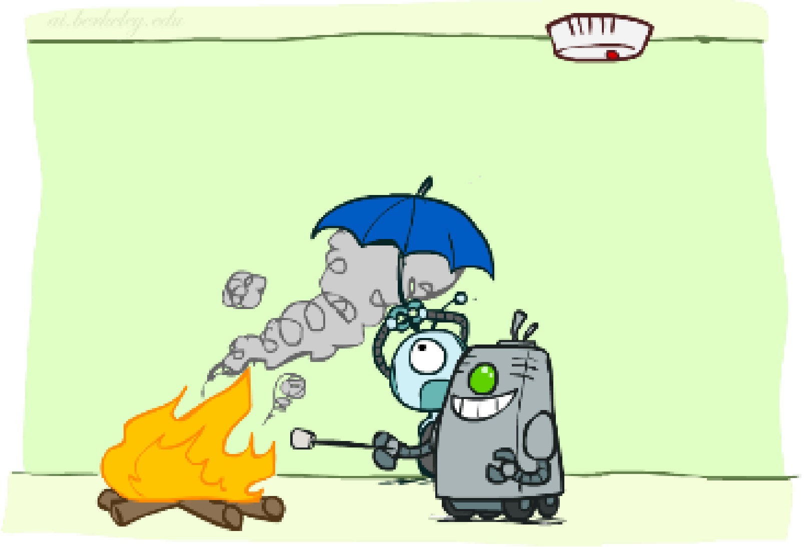


Example: Smoke alarm

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

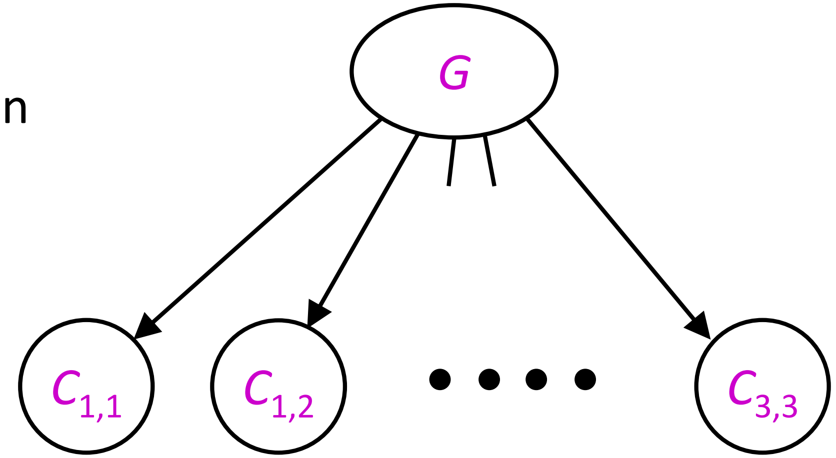


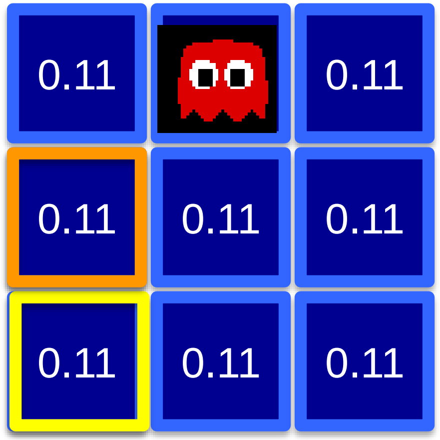




Example: Ghostbusters

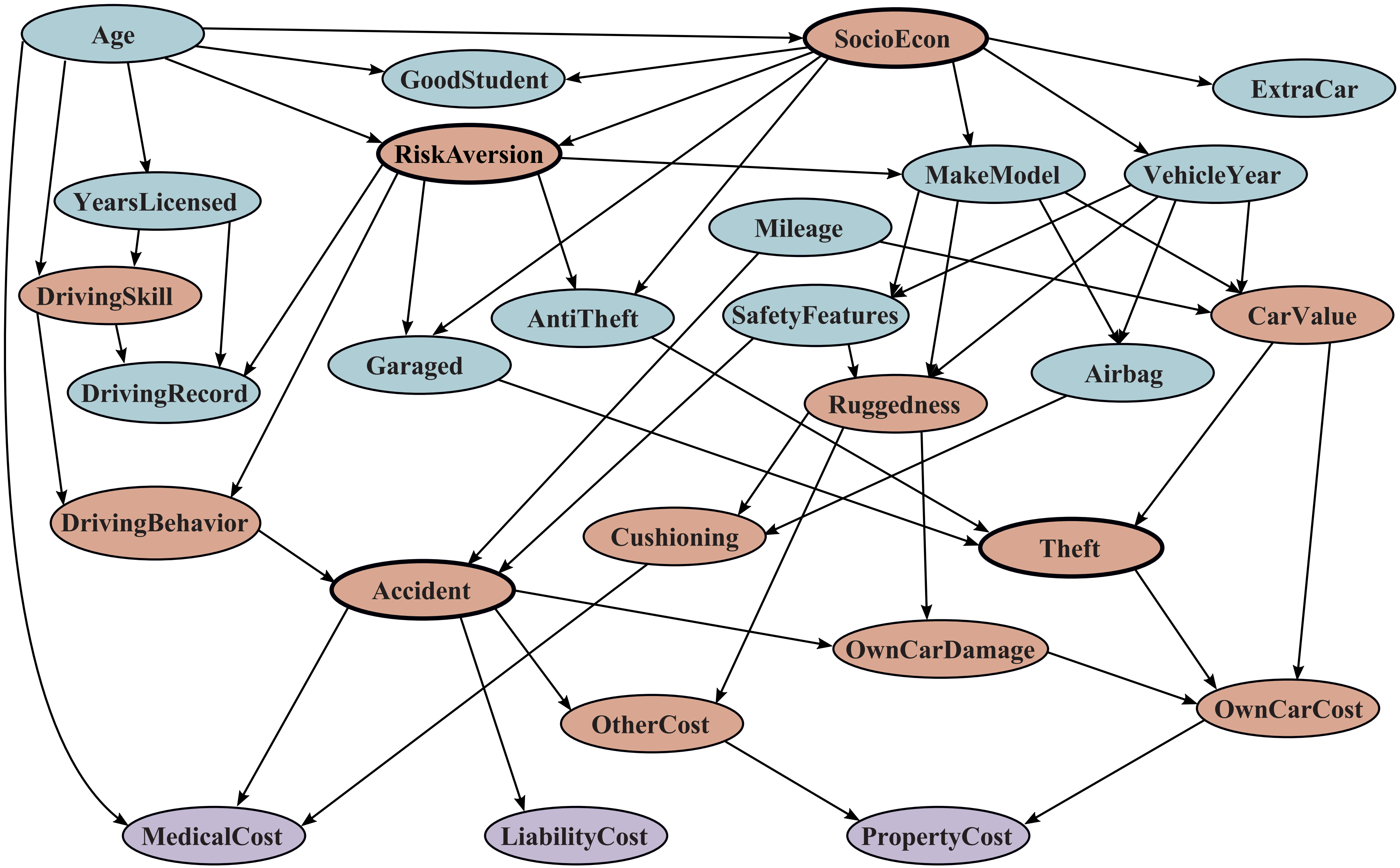
- Variables:
  - G: The ghost's location
  - $C_{1,1}$, … $C_{3,3}$:
    The observation at each location
- Want to estimate:
  $P(G \mid C_{1,1}, \ldots C_{3,3})$



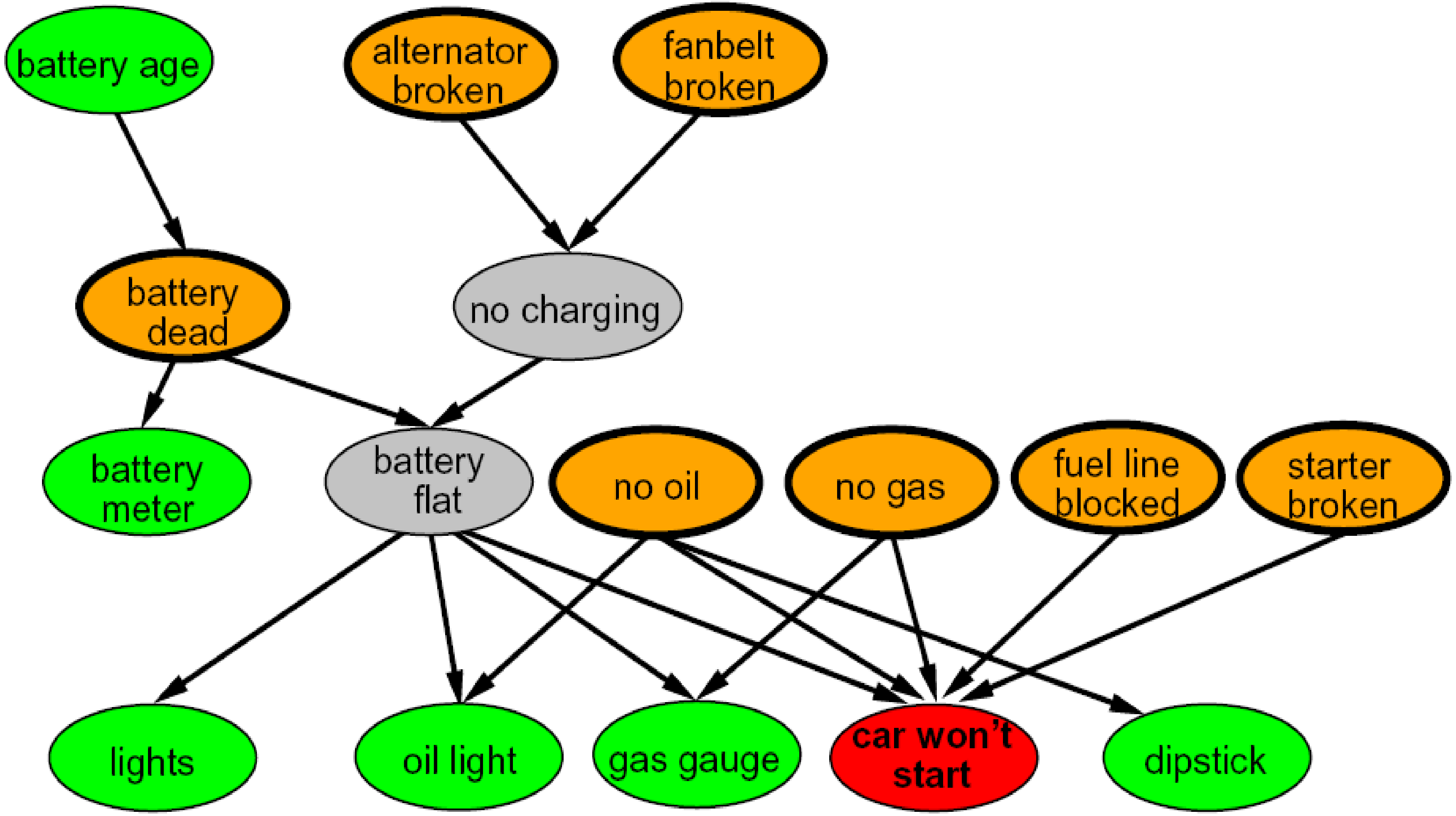


- This is called a ***Naïve Bayes*** model:
  - One discrete query variable (often called the ***class*** or ***category*** variable)
  - All other variables are (potentially) evidence variables
  - Evidence variables are all conditionally independent given the query variable

Example: Car Insurance



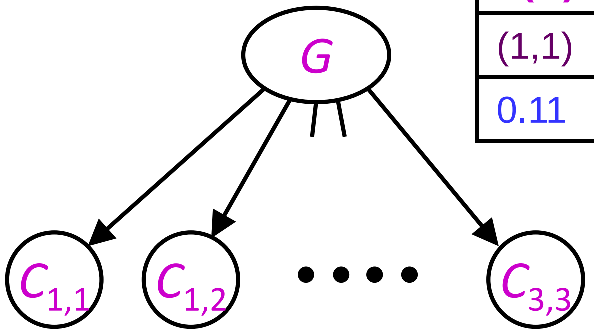
Example: Car Won’t Start



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 values of its parents



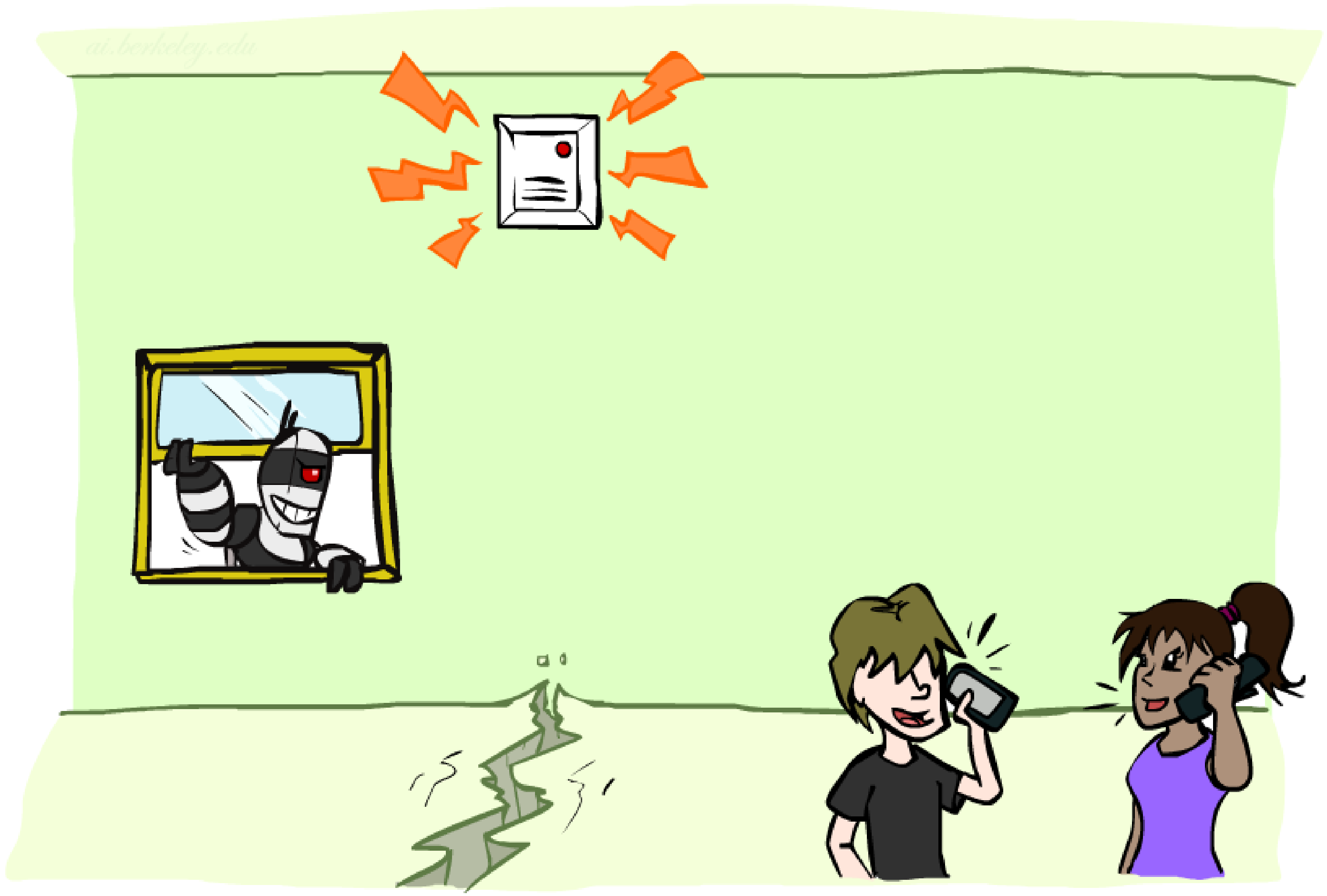
P(*G*)			
(1,1)	(1,2)	(1,3)	…
0.11	0.11	0.11	…

G	**P($C_{1,1}$ | G)**			
	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
…				

Bayes net = Topology (graph) + Local Conditional Probabilities

Example: Alarm Network

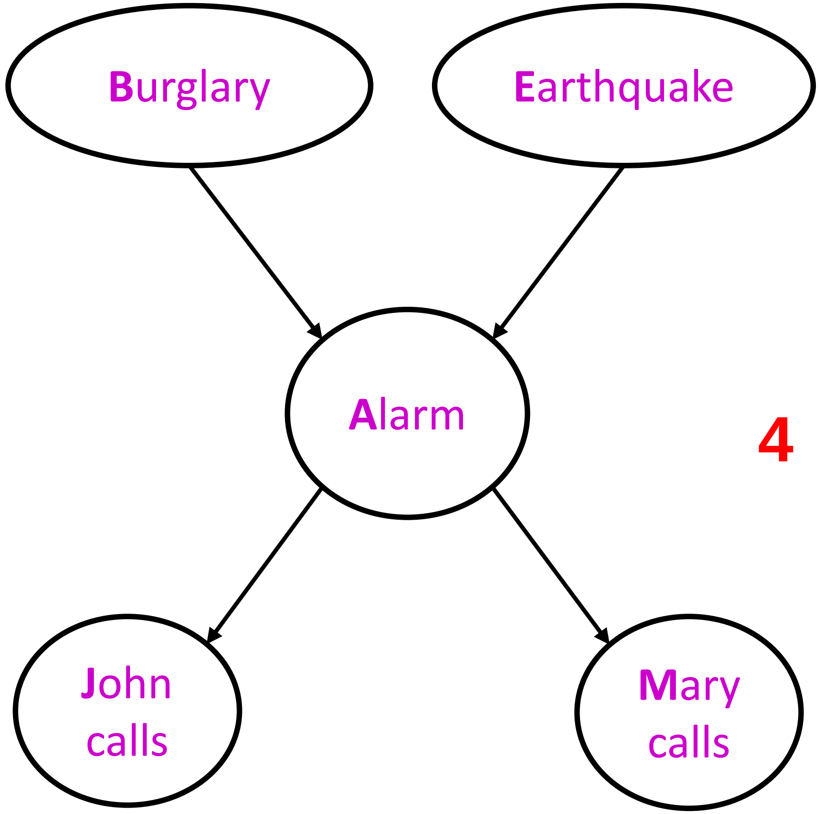
- Variables
  - B: Burglary
  - E: Earthquake
  - A: Alarm goes off
  - J: John calls
  - M: Mary calls



Example: Alarm Network

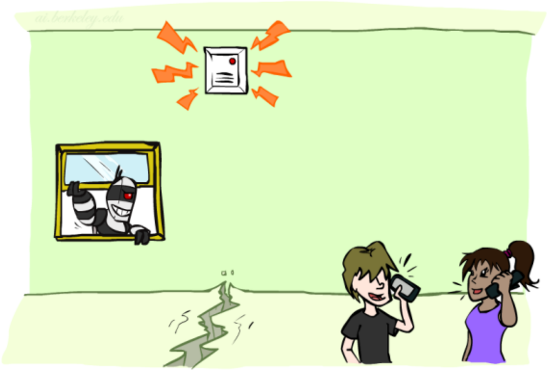
P(B)	
true	false
0.001	0.999

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

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

2

Number of ***free parameters*** in each CPT:

Parent range sizes $d_1,\ldots,d_k$

Child range size d
Each table row must sum to 1

$(d-1)\prod_i d_i$

Bayes net global semantics

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

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

- Exploits sparse structure: number of parents is usually small

Size of a Bayes Net

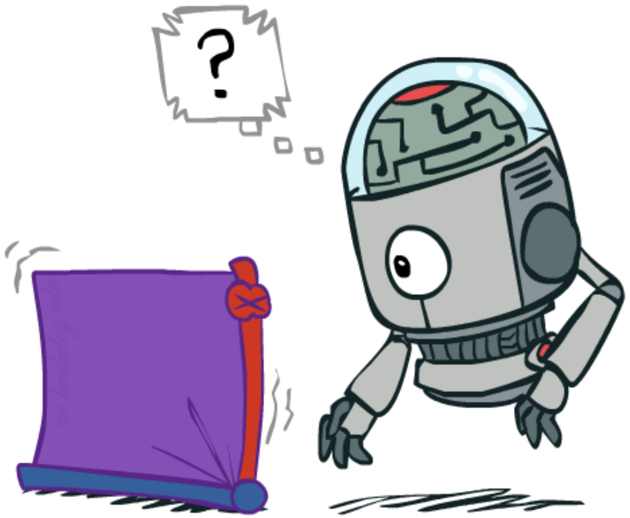
- How big is a joint distribution over *N* variables, each with *d* values?

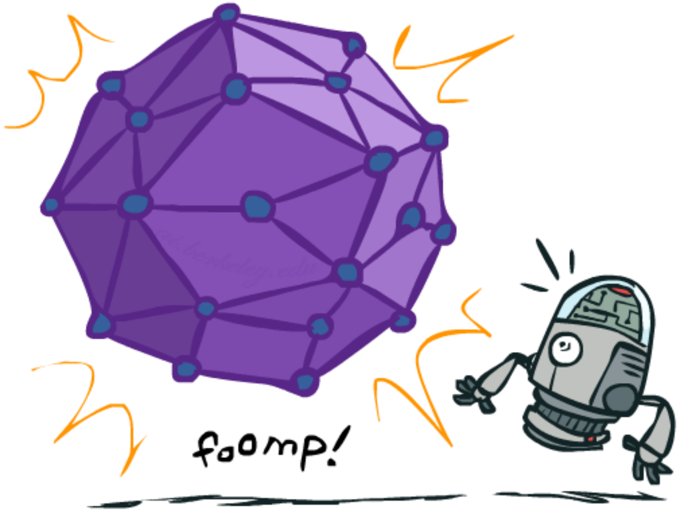
  d^N

- How big is an *N*-node net if nodes have at most *k* parents?

  $O(N * d^k)$

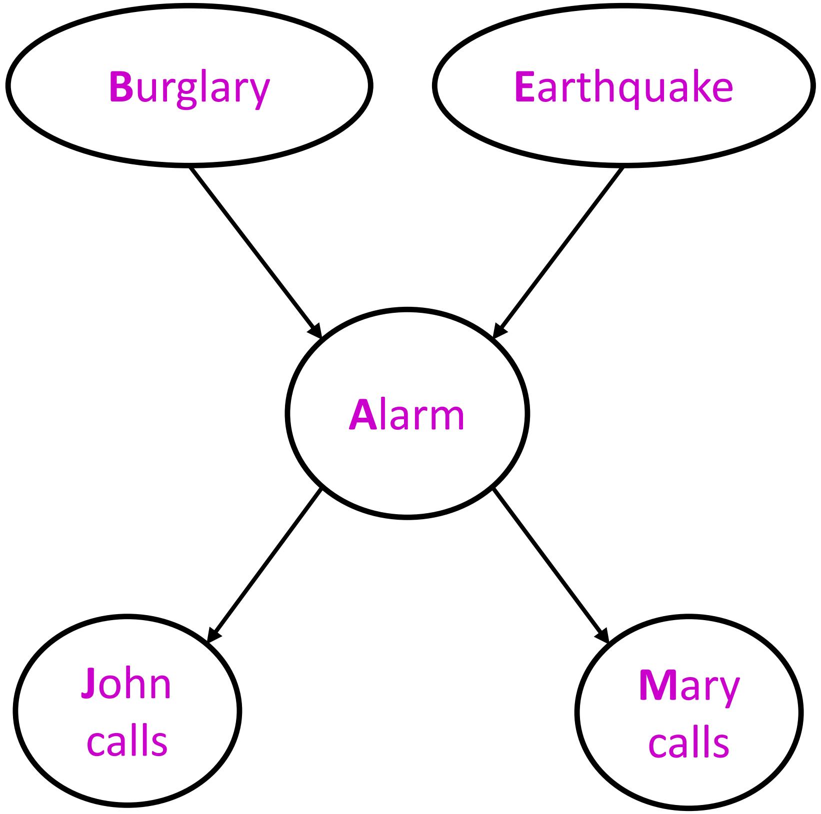
- Both give you the power to calculate $P(X_1, X_2, \ldots, X_N)$
- Bayes Nets: huge space savings with sparsity!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)





Example

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

$P(b, \neg e, a, \neg j, \neg m) =$

$P(b)\ P(\neg e)\ P(a|b,\neg e)\ P(\neg j|a)\ P(\neg m|a)$

$= .001 \times .998 \times .94 \times .1 \times .3 = .000028$

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

Conditional independence in BNs

- Compare the Bayes net global semantics

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

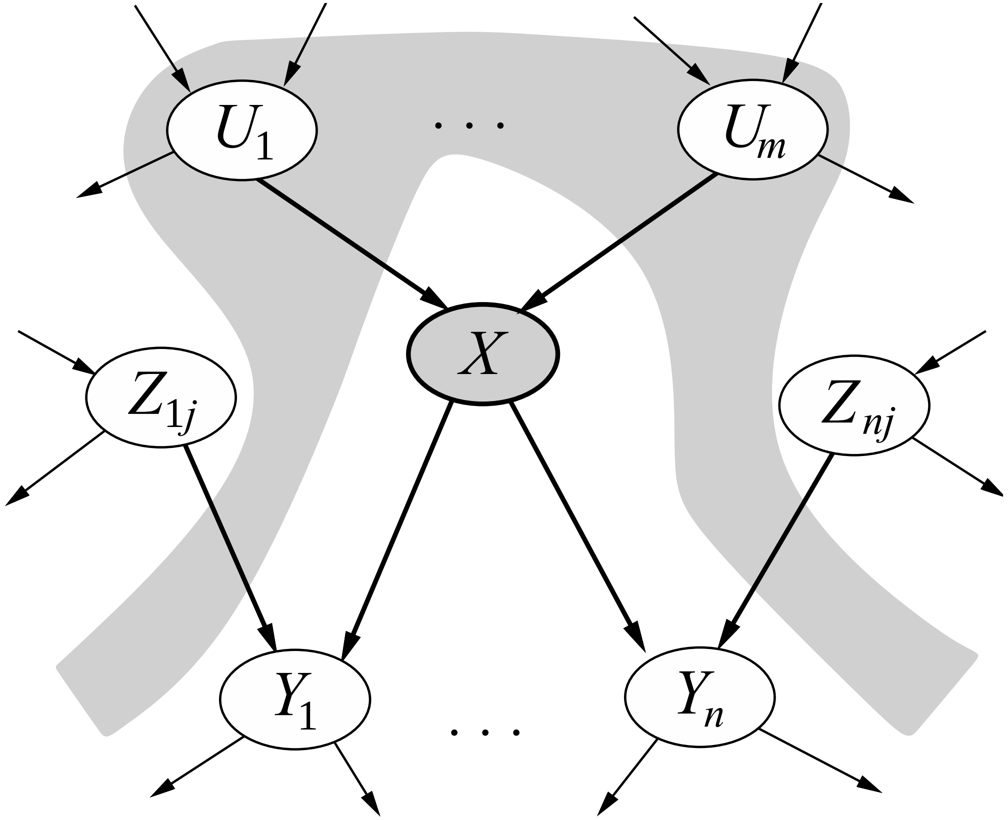
with the chain rule identity

$$P(X_1,..,X_n) = \prod_i P(X_i \mid X_1,\ldots,X_{i-1})$$

- Assume (without loss of generality) that $X_1,..,X_n$ sorted in topological order according to the graph (i.e., parents before children), so $Parents(X_i) \subseteq X_1,\ldots,X_{i-1}$
- So the Bayes net asserts conditional independences $P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid Parents(X_i))$
  - To ensure these are valid, choose parents for node X_i that "shield" it from other predecessors

Conditional independence semantics

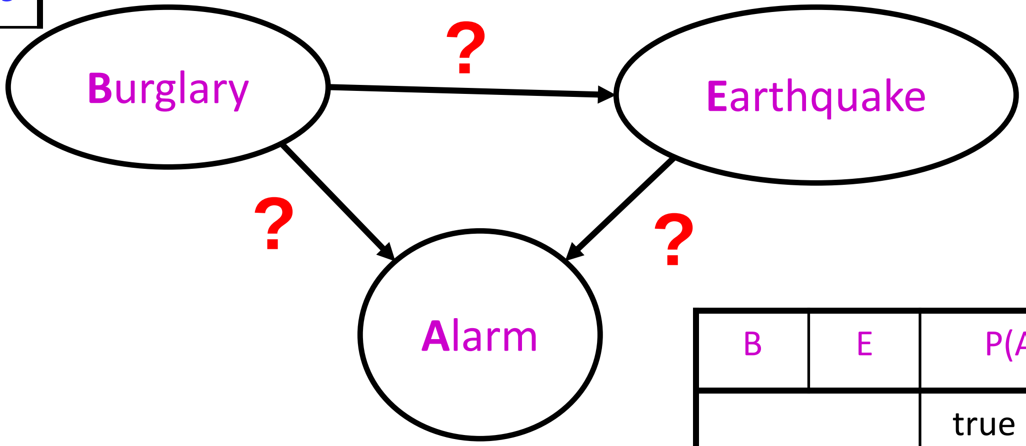
- ***Every variable is conditionally independent of its non-descendants given its parents***
- Conditional independence semantics <=> global semantics



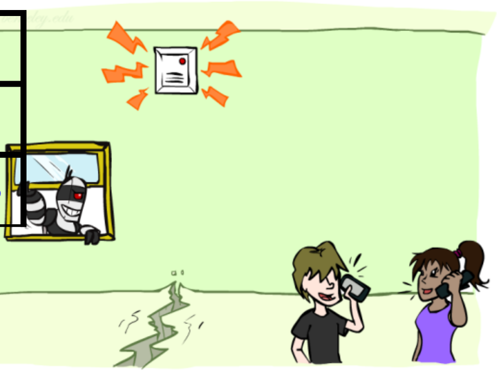
Example: Burglary

- Burglary
- Earthquake
- Alarm

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

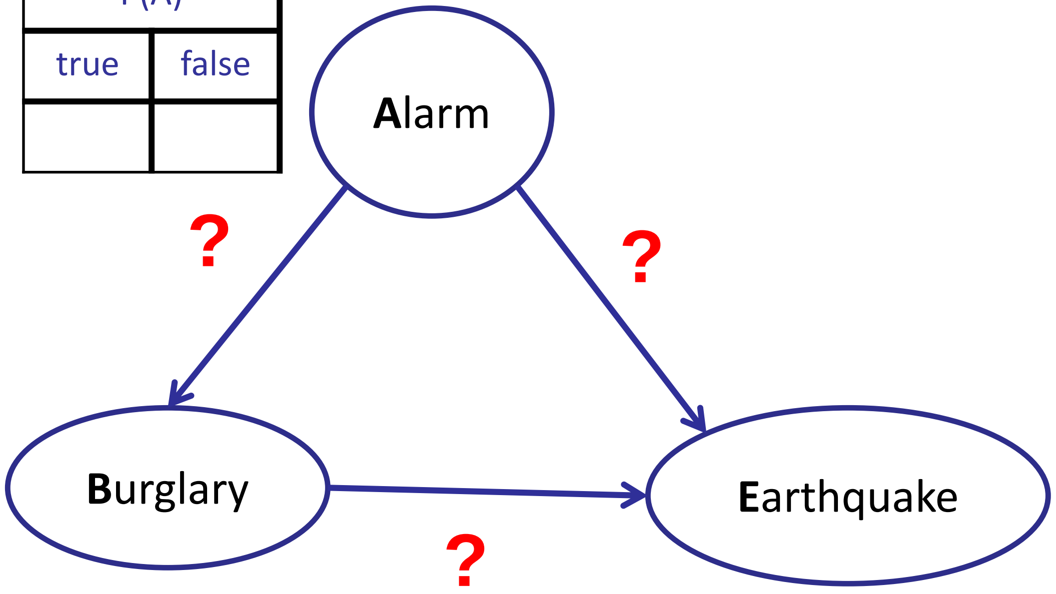


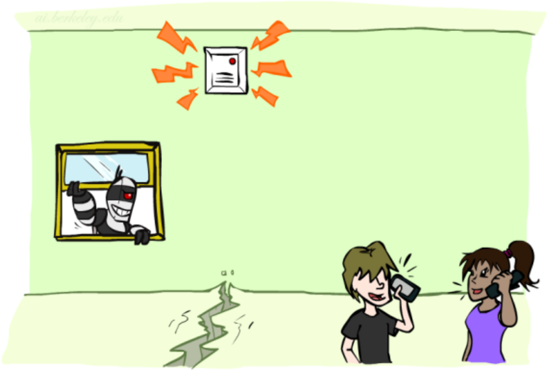
B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

Example: Burglary

- Alarm
- Burglary
- Earthquake

P(A)	
true	false

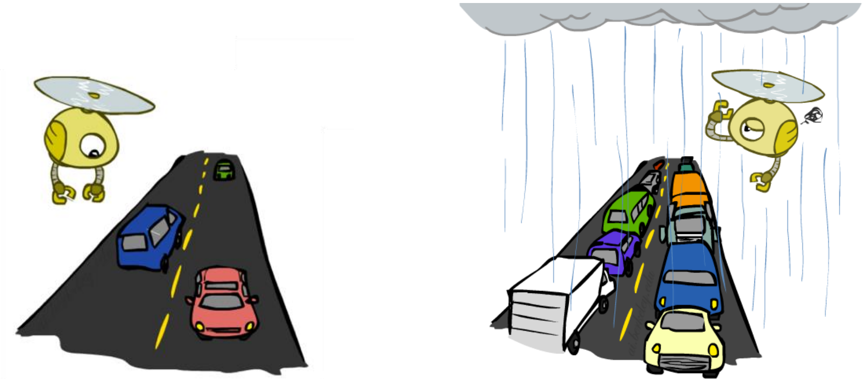


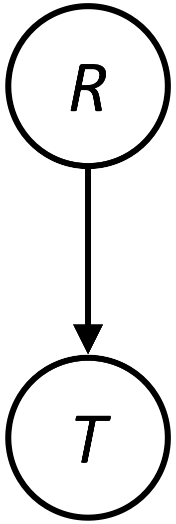


A	P(B|A)	
	true	false
true	?	
false		

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

Example: Traffic





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T, R)$

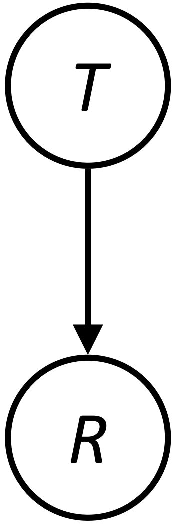
+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

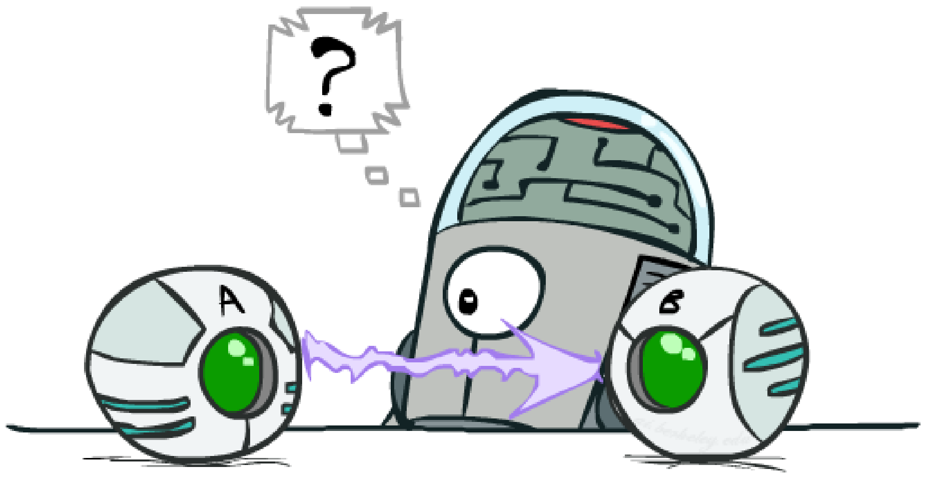
+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7



Causality?

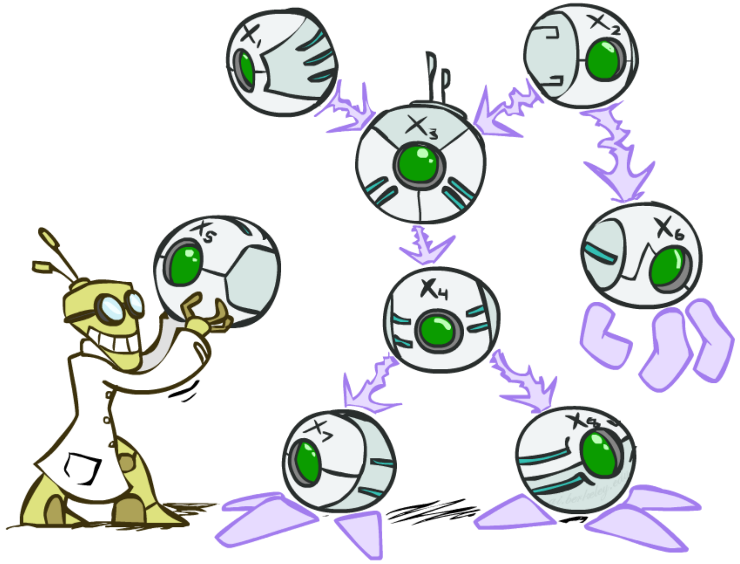
- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *Rain*
  - 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

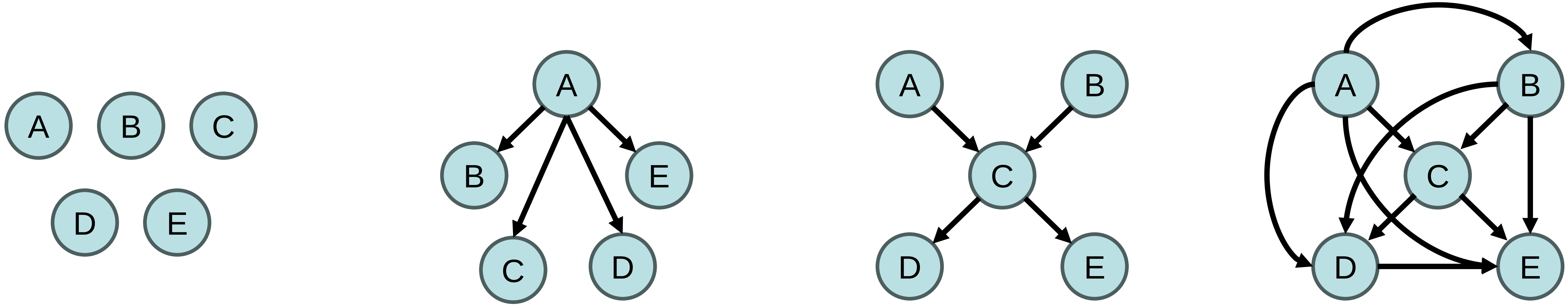
$$P(x_i|x_1,\ldots x_{i-1}) = P(x_i|\mathit{parents}(X_i))$$



Summary

- Independence and conditional independence are important forms of probabilistic knowledge
- Bayes net encode joint distributions efficiently by taking advantage of conditional independence
  - Global joint probability = product of local conditionals
- Allows for flexible tradeoff between model accuracy and memory/compute efficiency





Strict Independence Naïve Bayes Sparse Bayes Net Joint Distribution