Bayesian Networks

a.k.a. belief nets, bayes nets

CS182/CogSci110/Ling109
Spring 2007
Leon Barrett
lbarrett@eecs.berkeley.edu

Bayes Nets

- Representation of probabilistic information
  - reasoning with uncertainty
- Example tasks
  - Diagnose a disease from symptoms
  - Predict real-world information from noisy sensors
  - Process speech
  - Parse natural language

This lecture

- Basic probability
  - distributions
  - conditional distributions
  - Bayes' rule
- Bayes nets
  - representation
  - independence
  - algorithms
  - specific types of nets
    - Markov chains, HMMs

Probability

- Random Variables
  - Boolean/Discrete
    - True/false
    - Cloudy/rainy/sunny
    - e.g. die roll, coin flip
  - Continuous
    - [0,1] (i.e. 0.0 <= x <= 1.0)
    - e.g. thrown dart position, amount of rainfall

Unconditional Probability

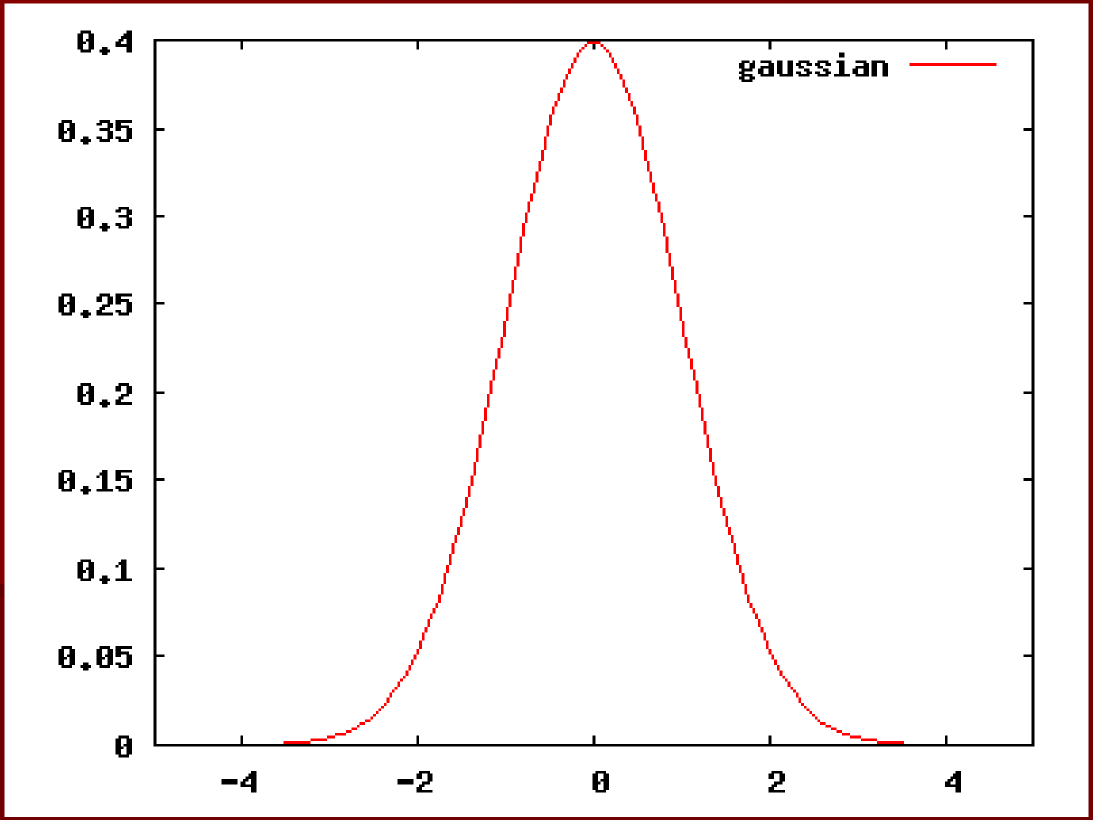
- Probability Distribution
  - In absence of any other info
  - Sums to 1
  - for discrete variable, it's a table
  - E.g. P(Sunny) = .65 (thus, P(¬Sunny) = .35)
  - for discrete variables, it's a table

Weather	sunny	cloudy	rainy	snowy
P(Weather)	0.65	0.19	0.14	0.02

Die	1	2	3	4	5	6
P(Die)	1/6	1/6	1/6	1/6	1/6	1/6

Continuous Probability

- Probability Density Function
  - Continuous variables
  - E.g. Uniform, Gaussian, Poisson…



Joint Probability

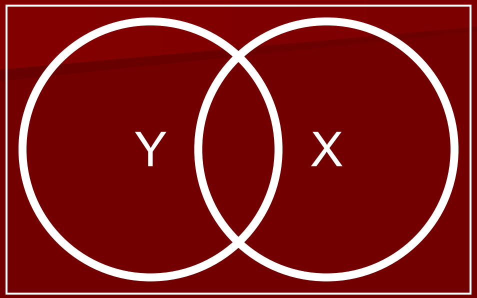
- Probability of several variables being set at the same time
  - e.g. P(Weather,Season)
- Still sums to 1
- 2-D table
- P(Weather, Season)

	sunny	cloudy	rainy	snowy	
summer	0.45	0.04	0.01	0	0.5
winter	0.2	0.15	0.13	0.02	0.5
	0.65	0.19	0.14	0.02	1

- Full Joint is a joint of all variables in model
- Can get “marginal” of one variable
  - sum over the ones we don't care about

Conditional Probability

- P(Y | X) is probability of Y given that all we know is the value of X
  - E.g. P(cavity | toothache) = .8
    - thus P(¬cavity | toothache) = .2
- Product Rule
  - P(X, Y) = P(Y | X) P(X)
  - P(Y | X) = P(X, Y) / P(X) (*normalizer* *to add up to 1*)



Conditional Probability Example

- P(disease=true) = 0.001 ; P(disease=false) = 0.999
- test 99% accurate:

P(test | disease)	true	false
positive	0.99	0.01
negative	0.01	0.99

- Compute joint probabilities
  - P(test=positive, disease=true) = 0.001 * 0.99 = 0.00099
  - P(test=positive, disease=false) = 0.999 * 0.01 = 0.00999
  - P(test=positive) = 0.00099 + 0.00999 = 0.01098

Bayes' Rule

- Result of product rule
  - $P(X, Y) = P(Y \mid X)\, P(X)$
    $= P(X \mid Y)\, P(Y)$

- $P(X \mid Y) = P(Y \mid X)\, P(X) / P(Y)$

- P(disease | test) = P(test | disease) * P(disease) / P(test)

Conditional Probability Example (Revisited)

- P(disease=true) = 0.001 ; P(disease=false) = 0.999
- test 99% accurate:

P(test | disease)	true	false
positive	0.99	0.01
negative	0.01	0.99

- P(disease=true | test=positive)
  = P(disease=true, test=positive) / P(test=positive)
- = 0.00099 / 0.01098 = 0.0901 = 9%

Important equations

- $P(X,Y) = P(X \mid Y)\,P(Y)$
  $= P(Y \mid X)\,P(X)$

- $P(Y \mid X) = P(X \mid Y)\,P(Y) / P(X)$

- Chain Rule of Probability

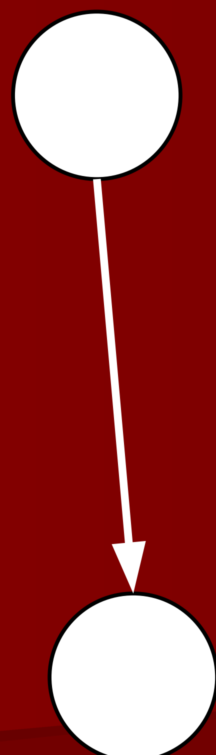
  $P(x_1,x_2,x_3,\ldots,x_k) =$

  $P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)\ldots P(x_k|x_1,x_2,\ldots,x_{k-1})$

  $P(x_1,x_2)$

  $P(x_1,x_2,x_3)$

Bayes Nets



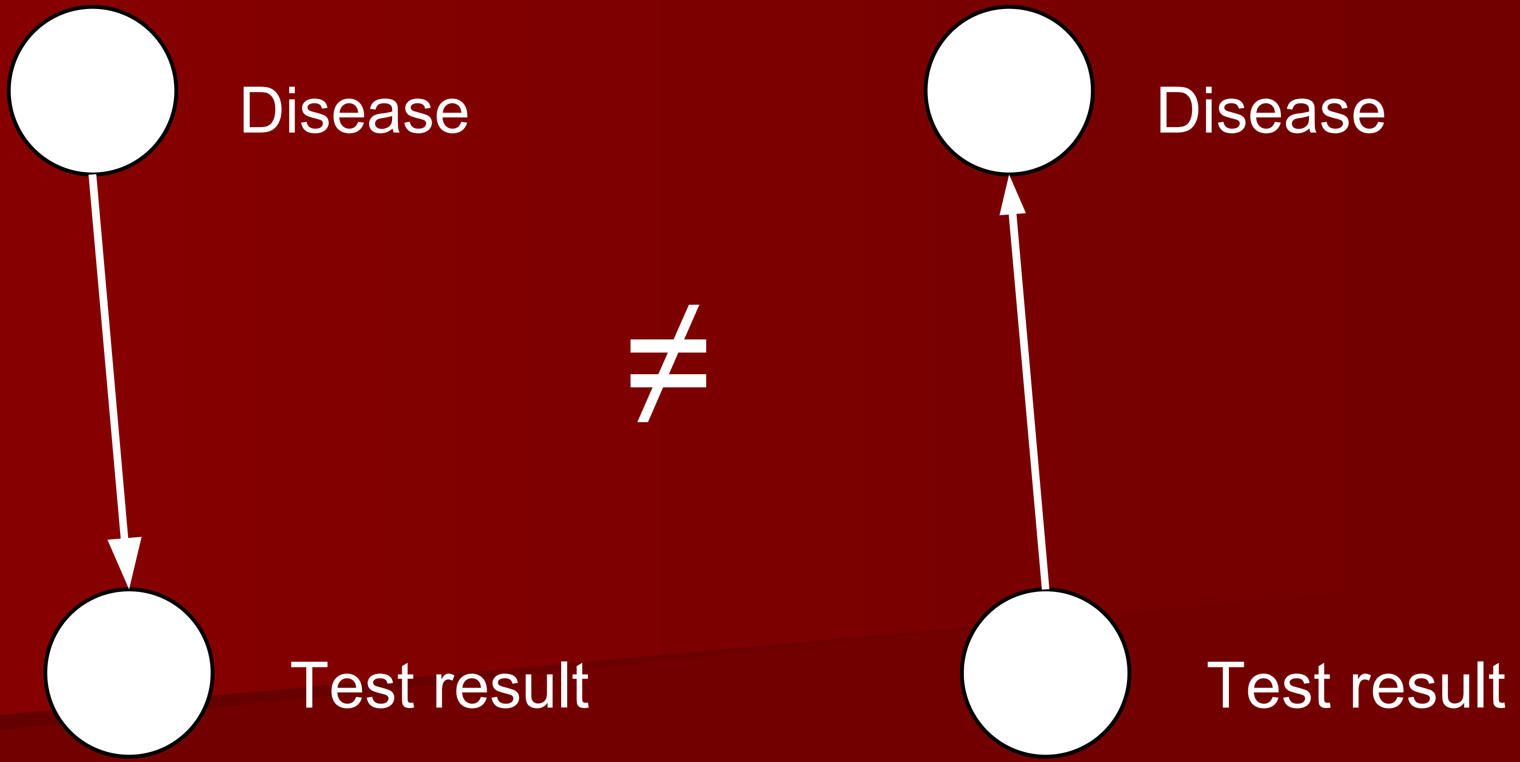
Disease

P(disease)	probability
TRUE	0.001
FALSE	0.999

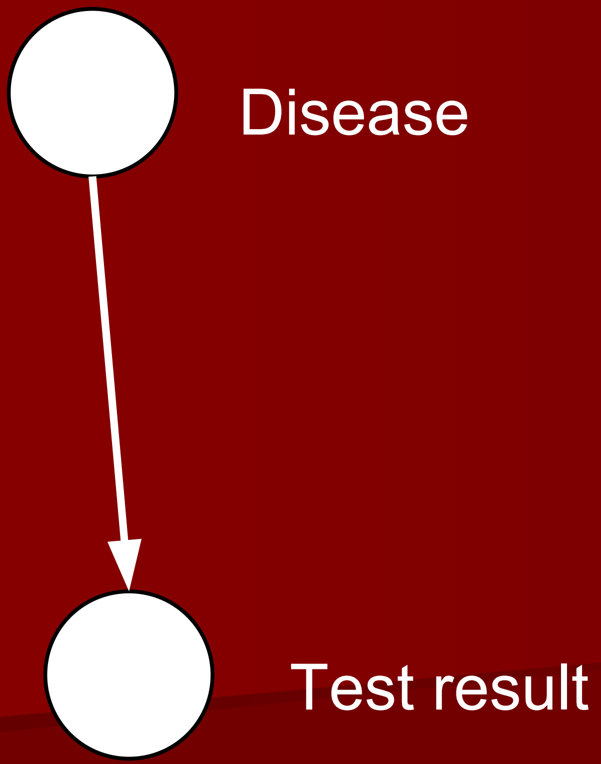
Test result

P(test | disease)	true	false
positive	0.99	0.01
negative	0.01	0.99

Causal reasoning

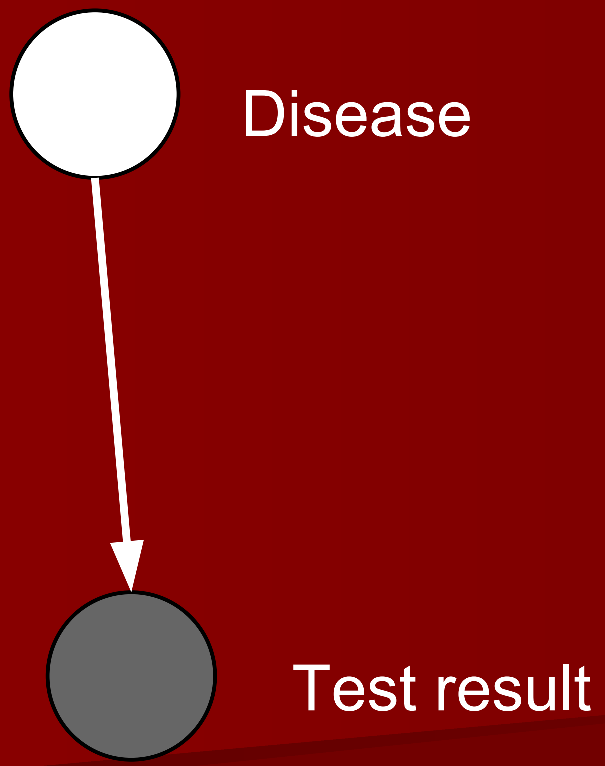


Causal reasoning



- not just probabilistic reasoning
- causal reasoning
  - arrow direction has important meaning
- manipulating causes changes outcomes
- manipulating outcomes does not change causes

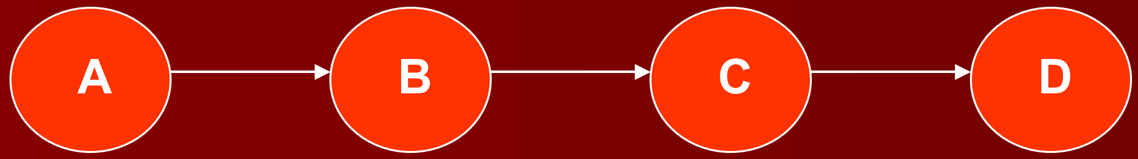
Bayes Nets



- Shaded means observed
  - we know the value of the variable
  - then we calculate P(net | observed)

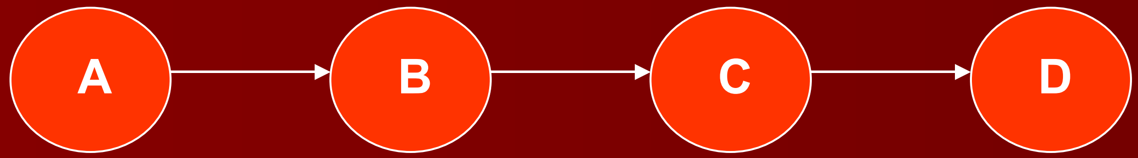
Example: Markov Chain

- Joint probability = P(A,B,C,D)
  = P(A)P(B|A)P(C|A,B)P(D|A,B,C) (by C.R.)



Example: Markov Chain

- Joint probability = P(A,B,C,D)
  = P(A)P(B|A)P(C|A,B)P(D|A,B,C) (by C.R.)

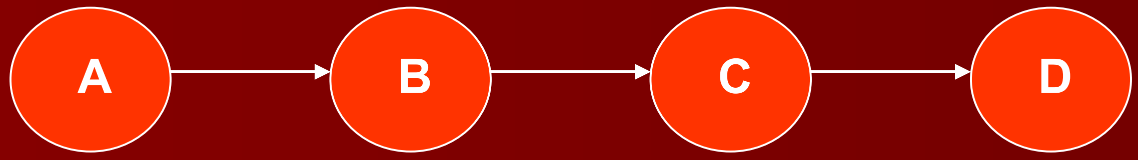


P(D|A,B,C) = P(D|C)

Example: Markov Chain

- Joint probability = P(A,B,C,D)

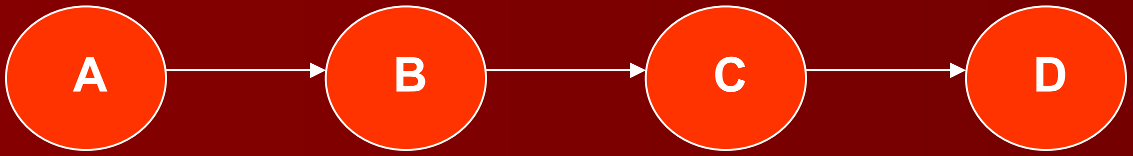
  = P(A)P(B|A)P(C|A,B)P(D|A,B,C) (by C.R.)



= P(A)P(B|A)P(C|B)P(D|C)

Example: Markov Chain

- Joint probability = P(A)P(B|A)P(C|B)P(D|C)



$$P(A) = \sum_B \sum_C \sum_D \hat{P}(A)\hat{P}(B|A)\hat{P}(C|B)\hat{P}(D|C)$$

$$P(A) = \hat{P}(A)\sum_B \hat{P}(B|A)\sum_C \hat{P}(C|B)\sum_D \hat{P}(D|C)$$

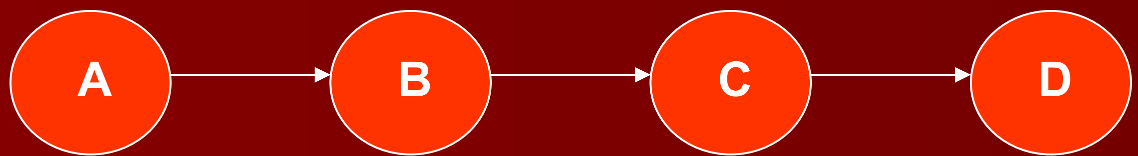
$$P(A) = \hat{P}(A)\sum_B \hat{P}(B|A)\sum_C \hat{P}(C|B)$$

$$P(A) = \hat{P}(A)\sum_B \hat{P}(B|A)$$

$$P(A) = \hat{P}(A)$$

Example: Markov Chain

- Joint probability = P(A)P(B|A)P(C|B)P(D|C)



$$P(C)=\sum_A\sum_B\sum_D \hat{P}(A)\hat{P}(B|A)\hat{P}(C|B)\hat{P}(D|C)$$

$$P(C)=\sum_B\sum_A \hat{P}(A)\hat{P}(B|A)\hat{P}(C|B)\sum_D \hat{P}(D|C)$$

$$P(C)=\sum_B \left[\sum_A \hat{P}(A)\hat{P}(B|A)\right] \hat{P}(C|B)$$

$$P(C)=\sum_B k(B)\hat{P}(C|B)$$

$$P(C)=k(C)$$

Variable Elimination

General idea:

- Write query in the form

$$P(X_n, \mathbf{e}) = \sum_{x_k} \cdots \sum_{x_3} \sum_{x_2} \prod_i P(x_i \mid pa_i)$$

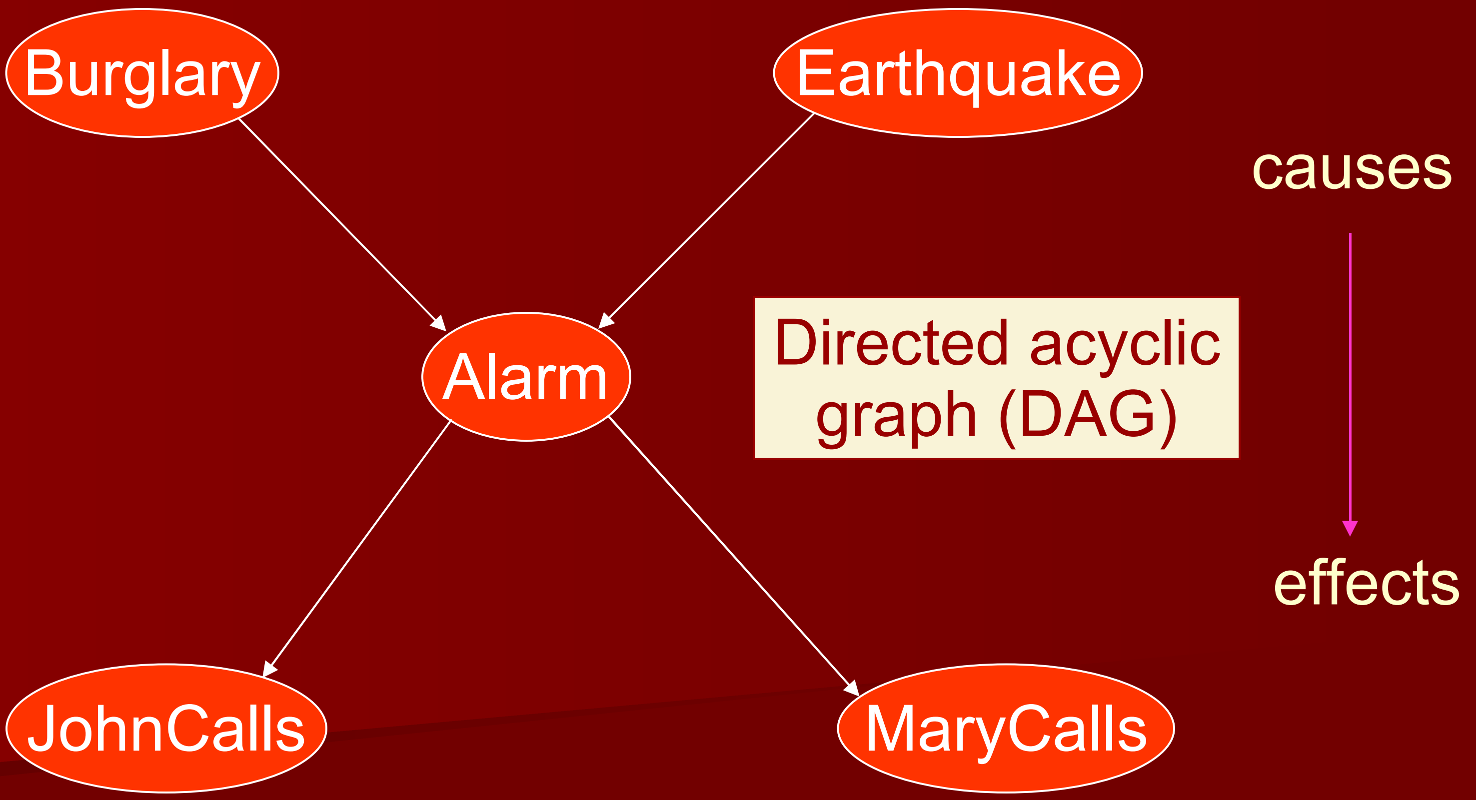
- Iteratively
  - Move all irrelevant terms outside of innermost sum
  - Perform innermost sum, getting a new term
  - Insert the new term into the product

Example: Alarm

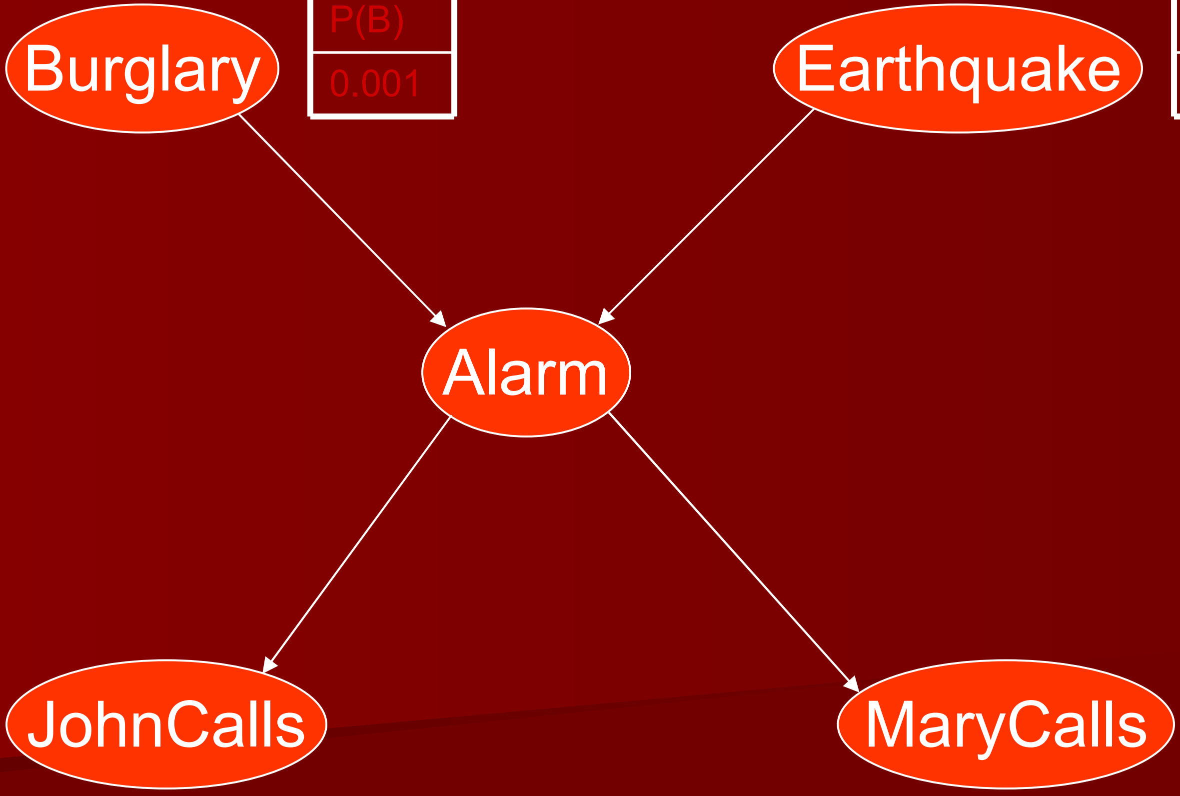
Five state features

- A: Alarm
- B: Burglary
- E: Earthquake
- J: JohnCalls
- M: MaryCalls

A Simple Bayes Net



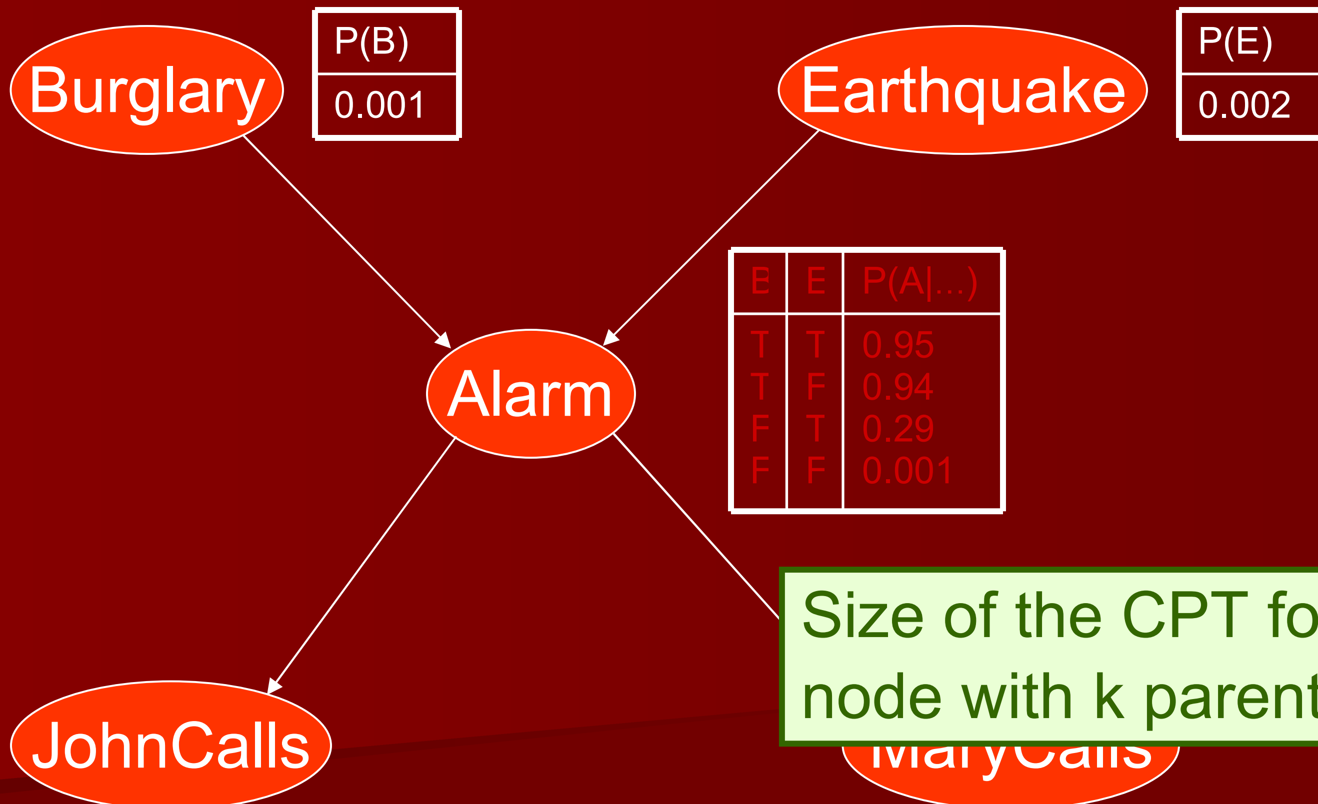
Assigning Probabilities to Roots



P(B)
0.001

P(E)
0.002

Conditional Probability Tables



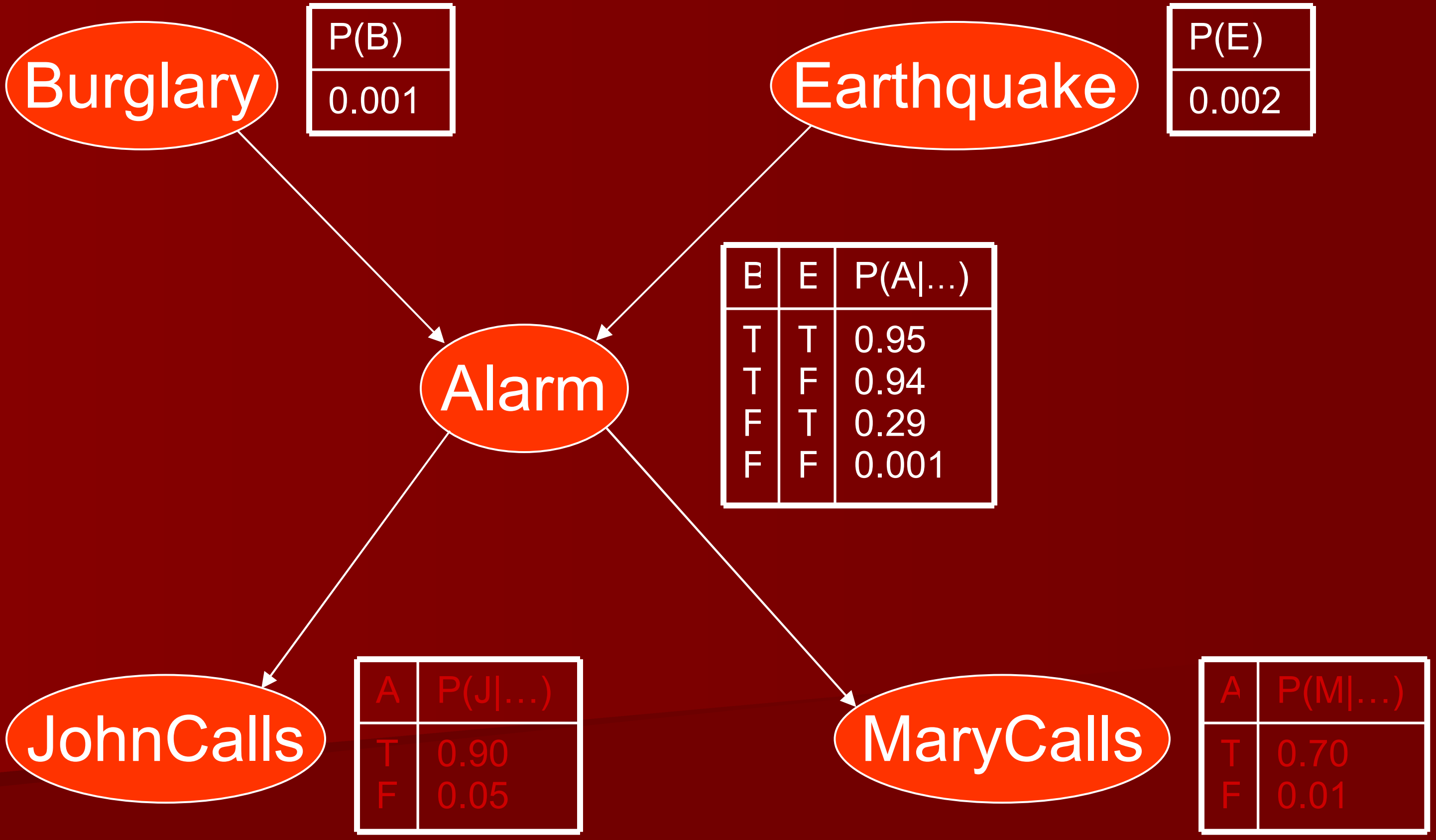
P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

Size of the CPT for a node with k parents: 2^{k+1}

Conditional Probability Tables



P(B)
0.001

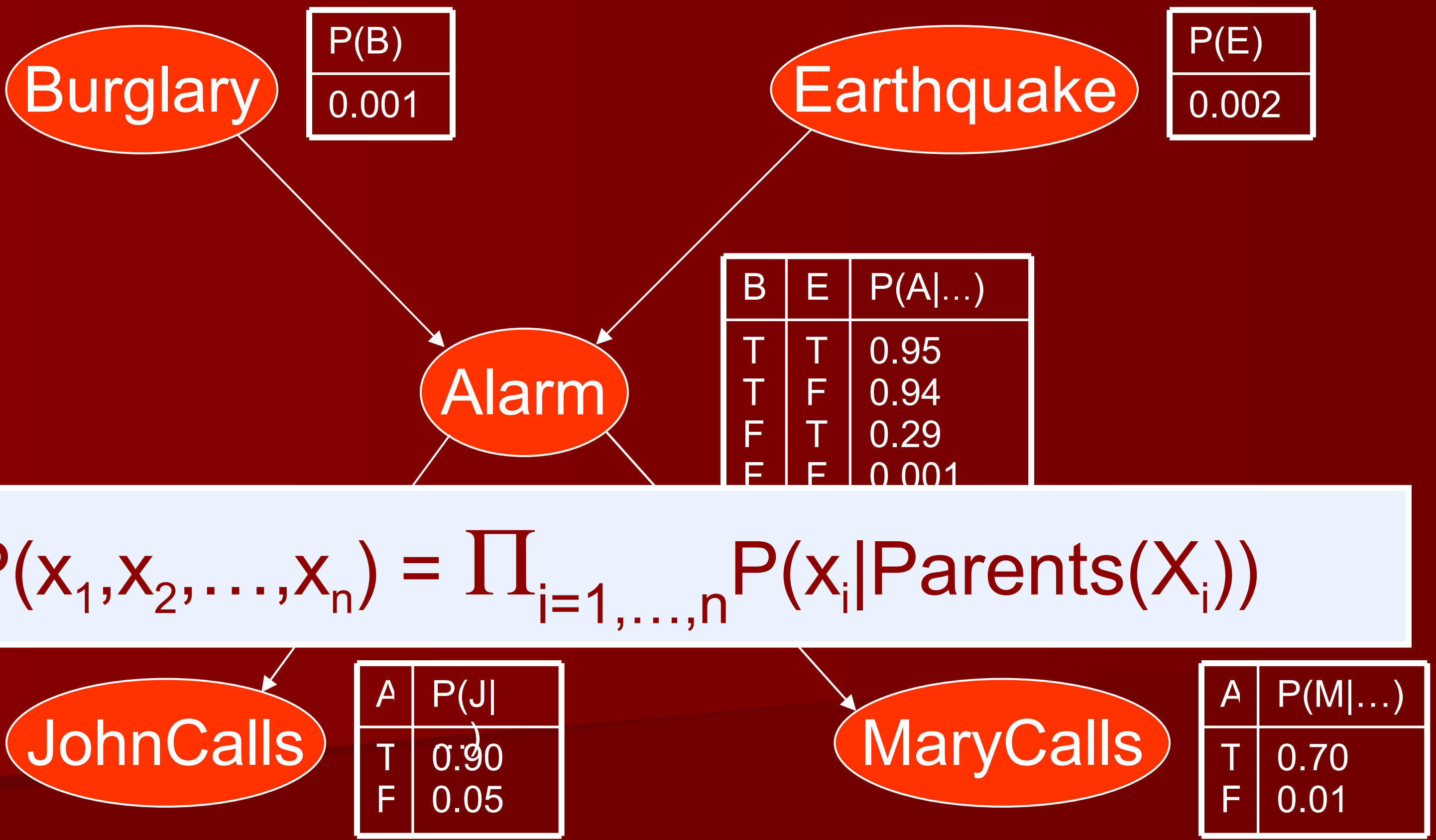
P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.90
F	0.05

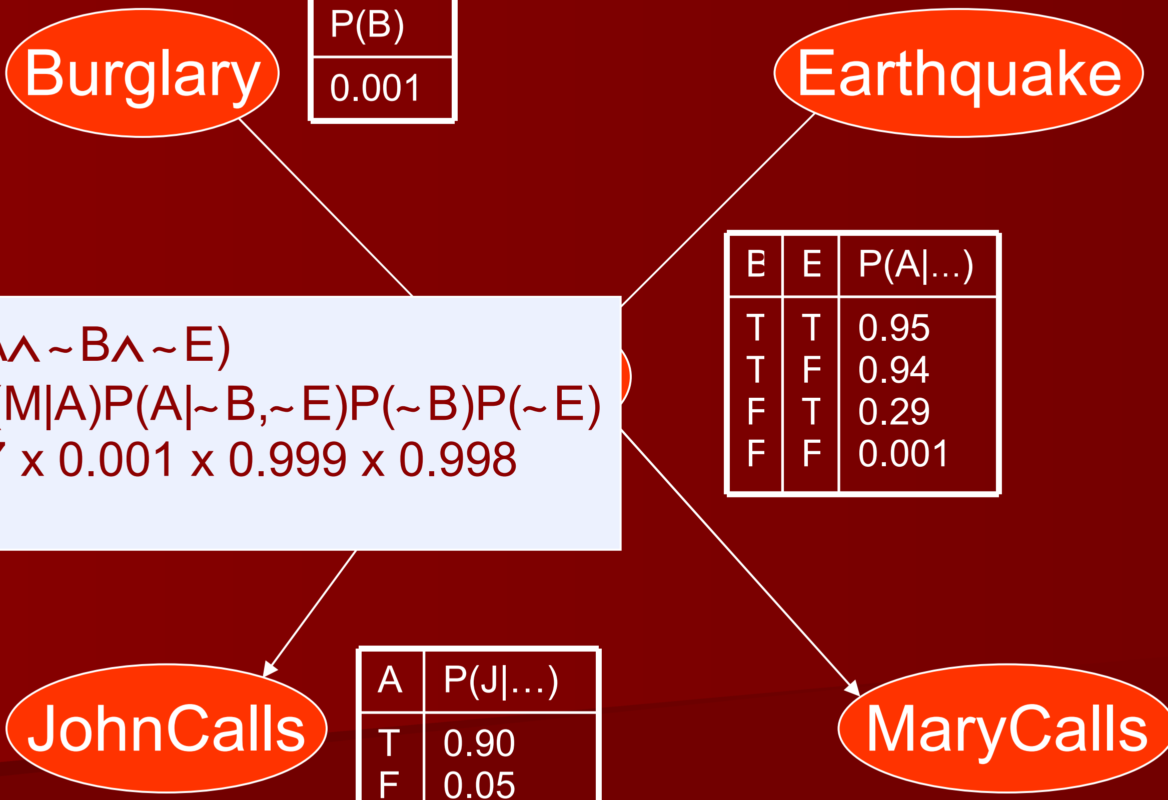
A	P(M|…)
T	0.70
F	0.01

What the BN Means



$$P(x_1,x_2,\ldots,x_n) = \prod_{i=1,\ldots,n} P(x_i|\text{Parents}(X_i))$$

Calculation of Joint Probability



P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.90
F	0.05

A	P(M|…)
T	0.70
F	0.01

$P(J \wedge M \wedge A \wedge \sim B \wedge \sim E)$
$= P(J|A)P(M|A)P(A|\sim B,\sim E)P(\sim B)P(\sim E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

Background: Independence

- Marginal independence:

$$X \perp Y := P(X,Y) = P(X)P(Y)$$

in other words,

$P(X|Y) = P(X)$ $P(Y|X) = P(Y)$

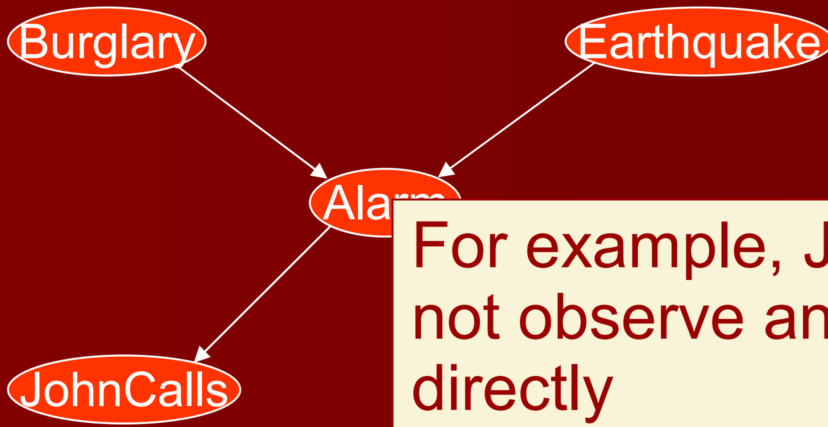
- Conditional independence

$$X \perp Y \mid Z := P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

or $:= P(X \mid Y, Z) = P(X \mid Z)$

Recall that $P(x|y) = P(x,y)/P(y)$

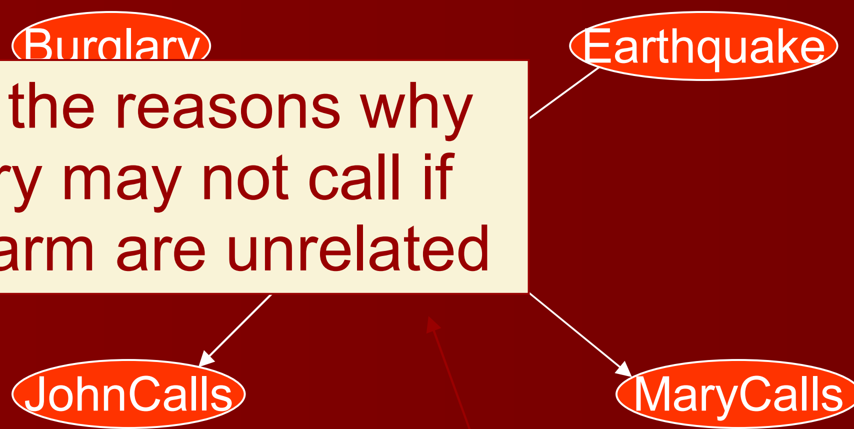
What the BN Encodes



For example, John does not observe any burglaries directly

- Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or ¬ Alarm
- The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬ Alarm

What the BN Encodes



For instance, the reasons why John and Mary may not call if there is an alarm are unrelated

- Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or ¬ Alarm
- The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬ Alarm

Independence

- Say we want to know the probability of some variable (e.g. JohnCalls) given evidence on another (e.g. Alarm). What variables are relevant to this calculation?
- I.e.: Given an arbitrary graph G = (V,E), is $X_A \perp X_B | X_C$ for some A,B, and C?
- The answer can be read directly off the graph, using a notion called D-separation

Independence

- Three cases:

Independence

- Three cases:

  (1) Markov Chain (linear)

$\sim(X_A \perp X_C)$

$X_A \perp X_c | X_B$

Independence

- Three cases:

  (2) Common Cause Model (diverging)

$\sim(X_A \perp X_C)$

$X_A \perp X_c | X_B$

Independence

- Three cases:

  (3) “Explaining away” (converging)

$X_A \perp X_C$

$\sim(X_A \perp X_c | X_B)$

Structure of BN

- The relation:

  $$P(x_1,x_2,\ldots,x_n) = \prod_{i=1,\ldots,n} P(x_i|Parents(X_i))$$

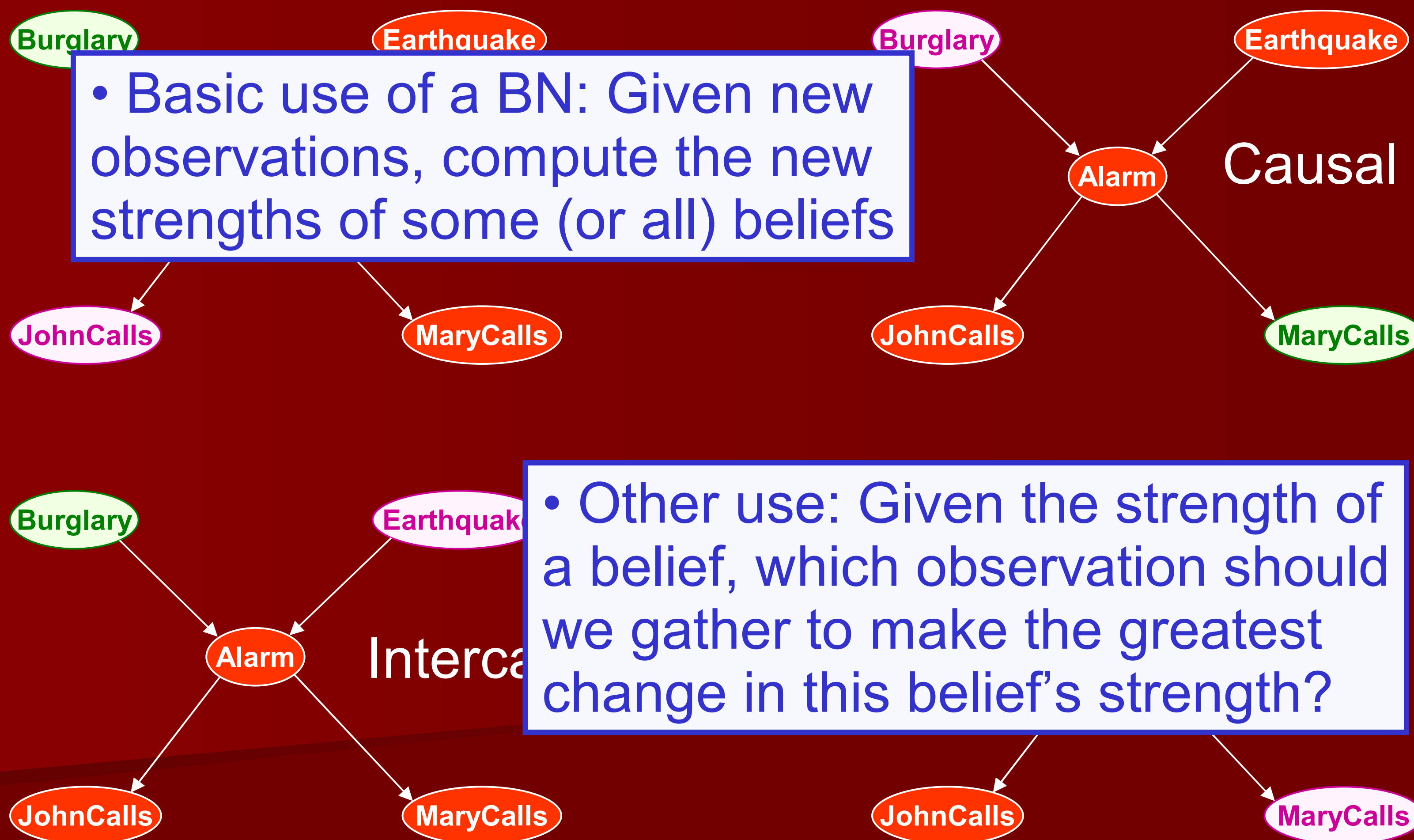
  means that each belief is independent of its predecessors in the BN given its parents
- Said otherwise, the parents of a belief X_i are all the beliefs that "directly influence" X_i

E.g., JohnCalls is influenced by Burglary, but not directly. JohnCalls is directly influenced by Alarm

Locally Structured Domain

- Size of CPT: 2^{k+1}, where k is the number of parents
- In a locally structured domain, each belief is directly influenced by relatively few other beliefs and k is small
- BN are better suited for locally structured domains

Inference Patterns



- Basic use of a BN: Given new observations, compute the new strengths of some (or all) beliefs

- Other use: Given the strength of a belief, which observation should we gather to make the greatest change in this belief's strength?

What can Bayes nets be used for?

- **Posterior probabilities**
  - Probability of any event given any evidence
- Most likely explanation
  - Scenario that explains evidence
- Rational decision making
  - Maximize expected utility
  - Value of Information
- Effect of intervention
  - Causal analysis

Explaining away effect

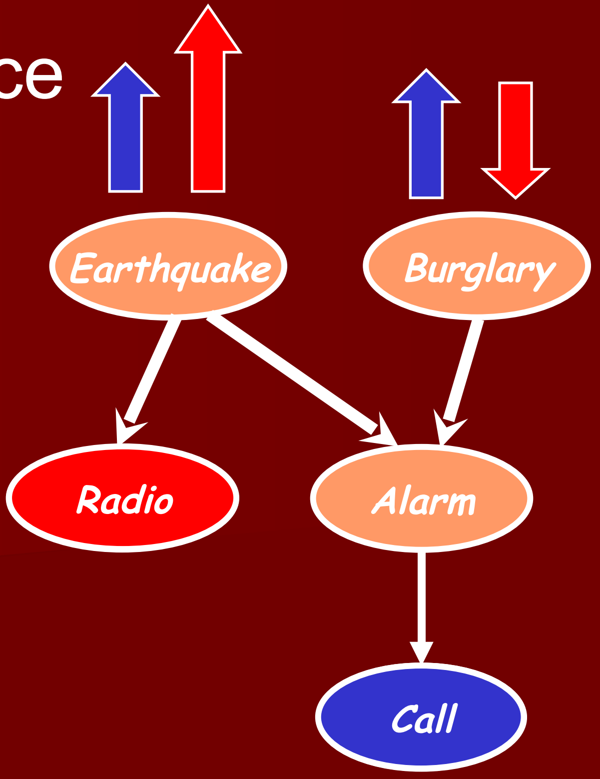
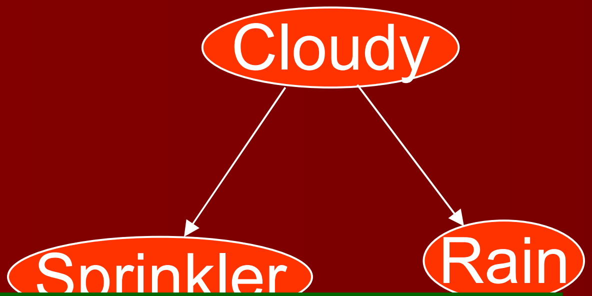


Figure from N. Friedman

Inference Ex. 2



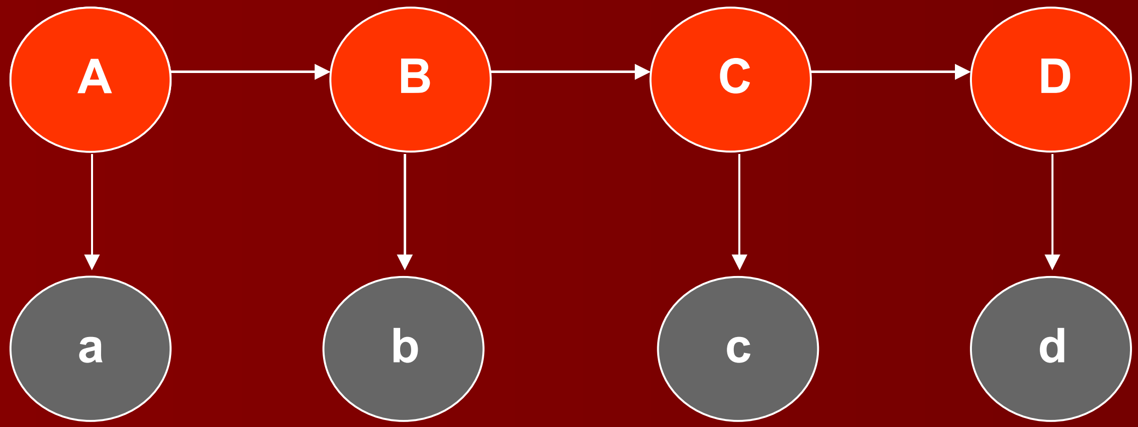
Algorithm is computing not individual

- Two ideas crucial to avoiding exponential blowup:
  - because of the structure of the BN, some subexpression in the joint depends only on a small number of variables
  - By computing them once and caching the result, we can avoid generating them exponentially many times

$$= \sum_{r,s} P(w \mid r,s) f_1(r,s)$$

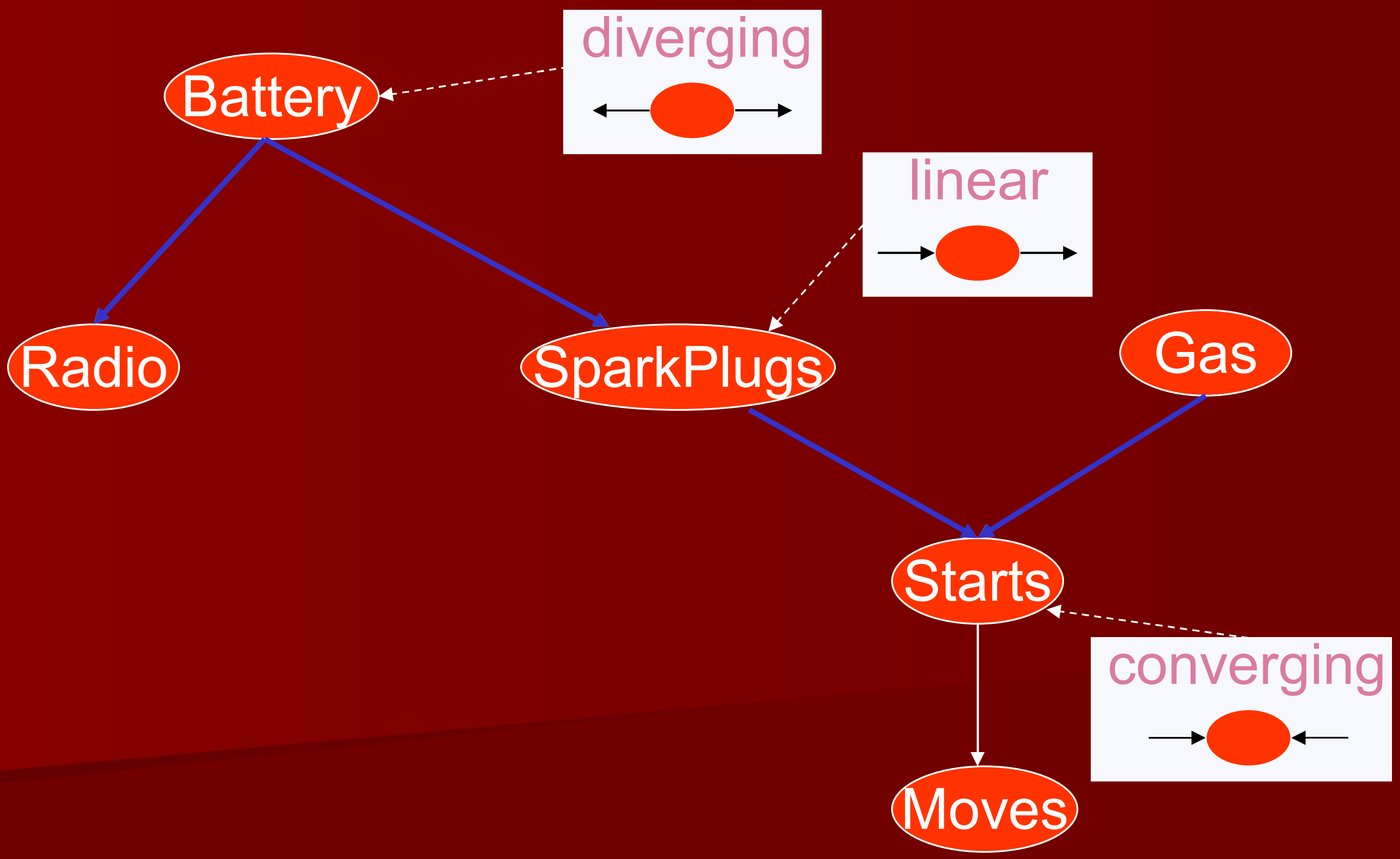
$f_1(r,s)$

Hidden Markov Models

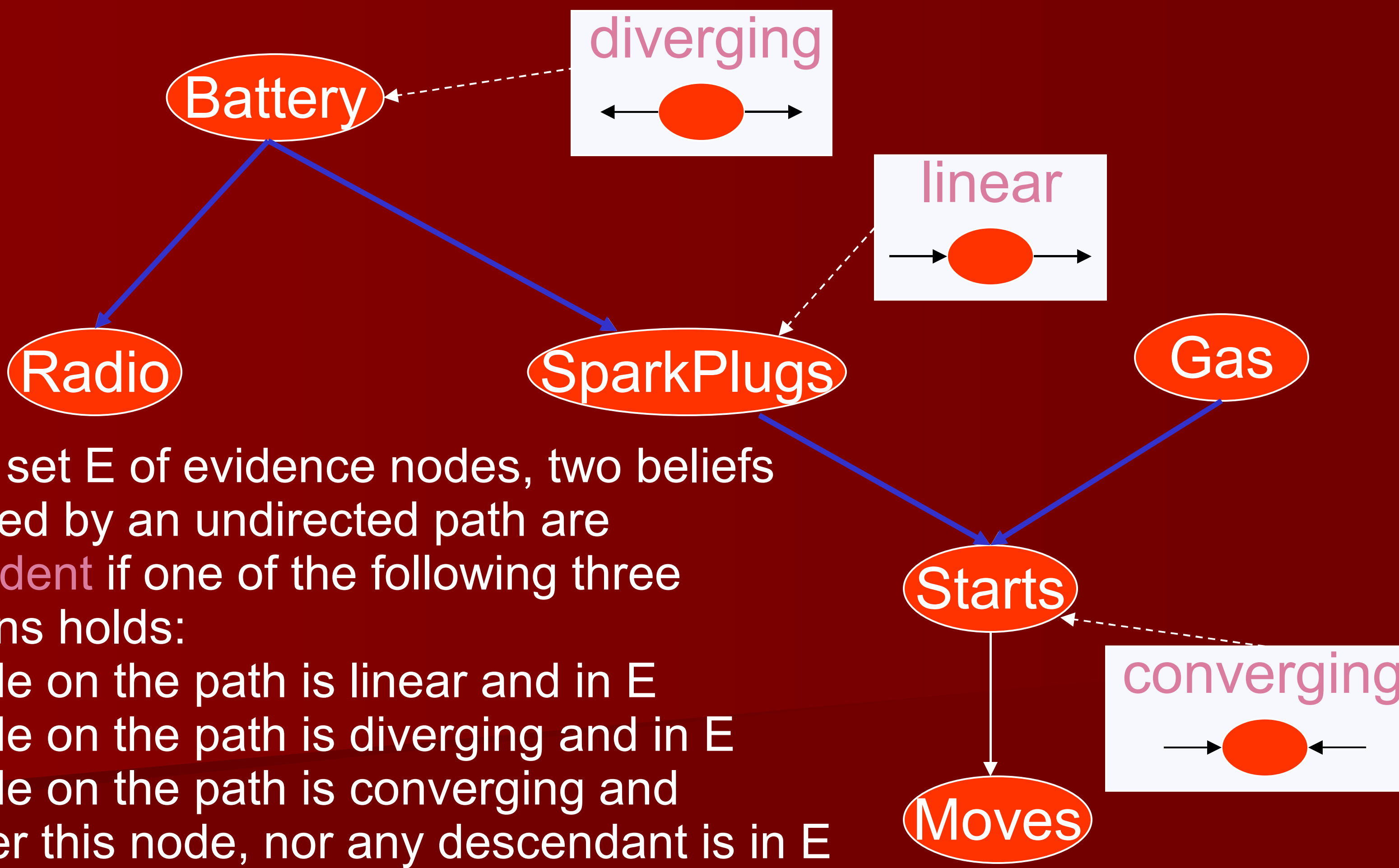


- Observe effects of hidden state
- Hidden state changes over time
- We have a model of how it changes
- E.g. speech recognition

Types Of Nodes On A Path



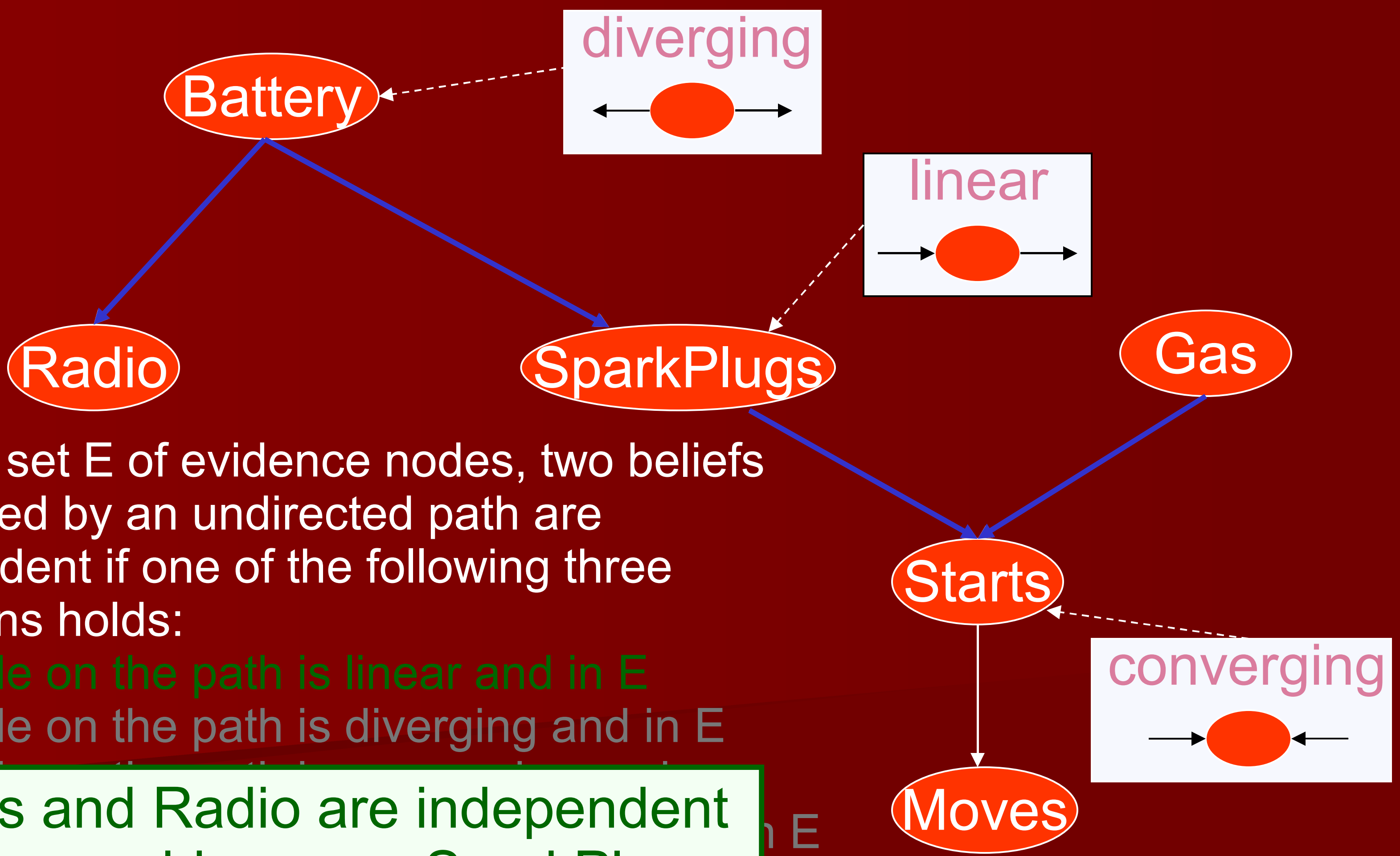
Independence Relations In BN



Given a set E of evidence nodes, two beliefs connected by an undirected path are independent if one of the following three conditions holds:

1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

Independence Relations In BN

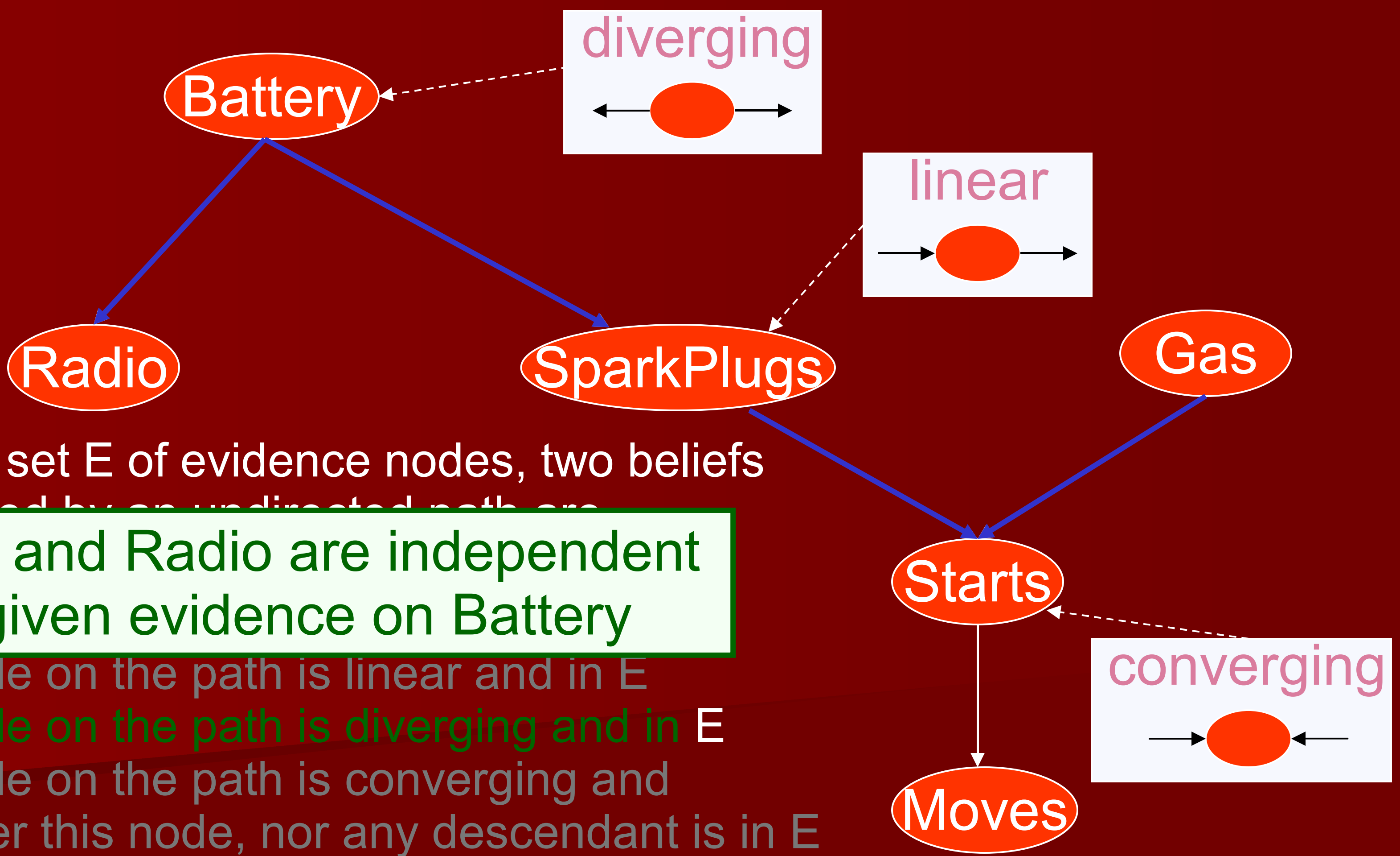


Given a set E of evidence nodes, two beliefs connected by an undirected path are independent if one of the following three conditions holds:

1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A … n … n E

Gas and Radio are independent given evidence on SparkPlugs

Independence Relations In BN

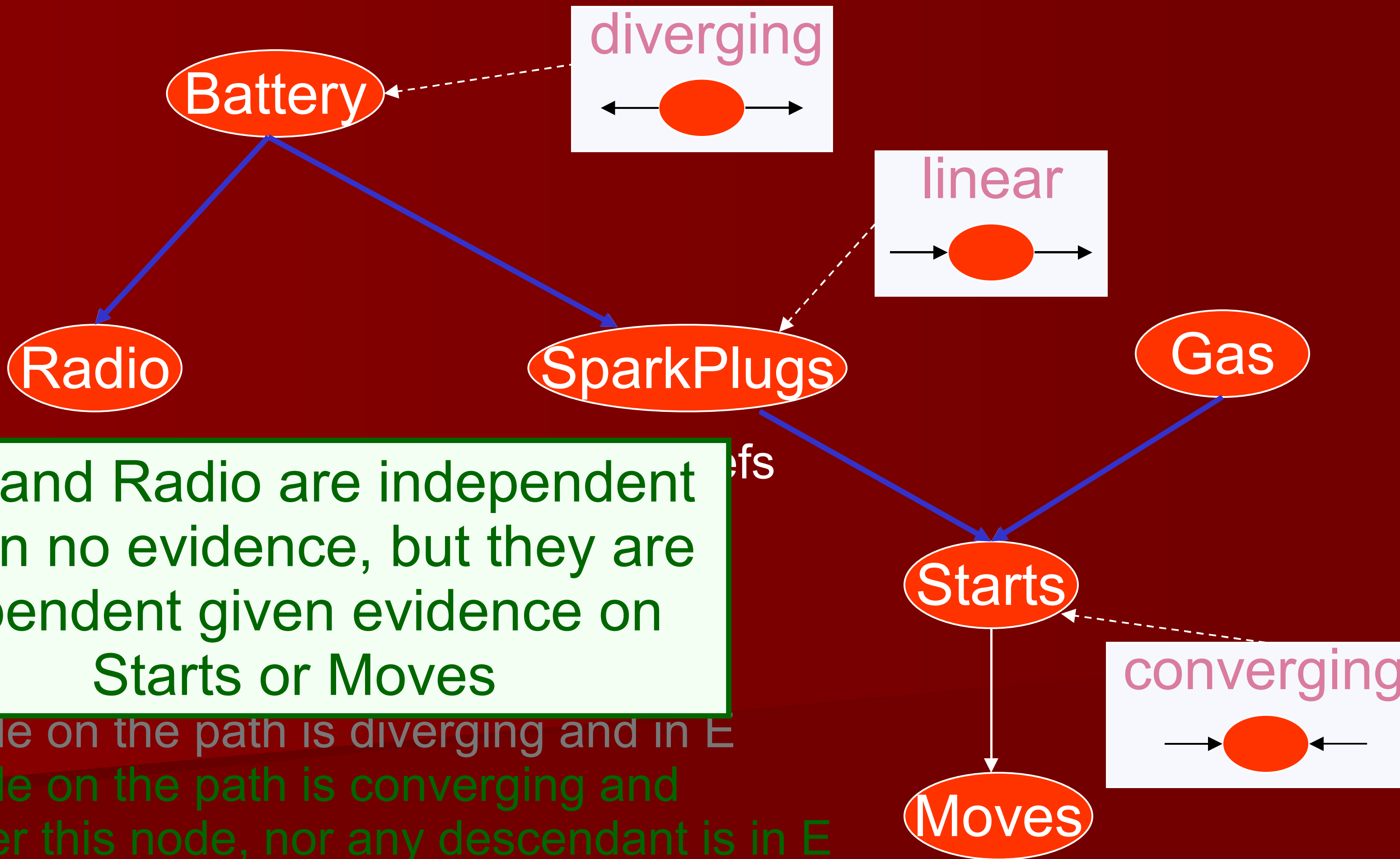


Given a set E of evidence nodes, two beliefs
co
in
co

Gas and Radio are independent given evidence on Battery

1. A node on the path is linear and in E
2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E

Independence Relations In BN



G … efs
c
in
c
1

Gas and Radio are independent given no evidence, but they are dependent given evidence on Starts or Moves

2. A node on the path is diverging and in E
3. A node on the path is converging and neither this node, nor any descendant is in E