Bayesian Networks

CS161
Prof. Guy Van den Broeck

Motivation from Logic

- Starting from logic:

$$Toothache \Rightarrow \; Cavity$$

- Rule is wrong; some exceptions:

$$Toothache \\ \Rightarrow \; Cavity \vee GumProblems \vee Abscess \vee \cdots$$

- This is intractable to model
- Perhaps make the rule causal:

$$Cavity \Rightarrow Toothache$$

… but this rule is wrong as well…

Motivation from Logic

The monotonicity of logic is the problem… again

- Either you model everything exhaustively
  – Intractable to model; we are too ignorant
  – Even if you could, you will never be able to act
- Assume too much and get stuck when things don't go as expected

Epistemological change: we no longer believe in a set of possible worlds (models), we believe in a probability for each world!

World View

- Propositional

  Global properties that are true or false
- Probabilistic
  - Belief is still a set of possible world
  - But now they have a degree of belief Pr(.)
  - Knowledge Base KB ≈ Pr
- *"Uncertainty is epistemological – pertaining to an agent's beliefs of the world – rather than ontological – how the world is."* [Poole et al.]
  - We can have different beliefs about the same world
  - What's the probability that the world ends tomorrow?

Decision-Making Motivation

- Acting with partial (noisy) sensor information

  $\Rightarrow$ Consider ever logically possible explanation
- Example: drive to airport

  $\Rightarrow$ No plan is guaranteed to achieve goal

  $\Rightarrow$ Yet the agent must act
- Decisions depend on
  - Relative importance of goals (*utility*)
  - The likelihood of achieving them (*probability*)

  $\Rightarrow$ Maximum expected utility

Propositions are only Boolean?

- Categorical variables
  - Weather=sunny, Weather=rainy, Weather=snowy
  - 3 Boolean variables that are mutually exclusive
    - Sometimes called “indicator variables”
    - Can all be encoded in sentences…
- Continuous variables
  - Temperature=73.514, Temperature=78.785, …
  - Infinitely many Boolean variables (and worlds).
    - In logic, see SAT Modulo Theories (SMT)
    - Special accommodations for continuous variables in statistics; we will mostly stick to the discrete world.

Sentences or “Events”

- Knowledge is a probability for every world: $\Pr(\omega)$
- What is the probability of a sentence α?
  (also called an “event” α in probability)
- Need to <u>axiomatize</u> probability [Kolmogorov]:
  1. Probabilities are non-negative: $0 \leq \Pr(\alpha)$
  2. The probability of a true event is 1: $\Pr(true) = 1$
  3. If α and β are mutually exclusive, then
     $\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta)$.

Sentences or "Events"

- Knowledge is a probability for every world: $\Pr(\omega)$
- What is the probability of a sentence α?
  (also called an "event" α in probability)
- A sentence α is equivalent to the disjunction of its models: $\alpha \equiv \omega_1 \vee \omega_8 \vee \omega_{11} \vee \omega_{17} \vee \cdots$

$$\boxed{\Pr(\alpha) = \sum_{\omega \models \alpha} \Pr(\omega) = \sum_{\omega \in Mods(\alpha)} \Pr(\omega)}$$

Properties of Probability

- Complement events
  - $\Pr(\alpha) + \Pr(\neg\alpha) = 1$
  - Why?
- Inclusion-exclusion
  - $\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha \wedge \beta)$
  - Why?

$$
\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\text{Earthquake} \wedge \text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) = .02 \\
\Pr(\text{Earthquake} \vee \text{Burglary}) &= .1 + .2 - .02 = .28
\end{aligned}
$$

[Darwiche ‘08]

Conditional Probability

- What if I observe new information in the form of a sentence β?
- Belief changes from $\Pr(\alpha)$ to $\Pr(\alpha|\beta)$
- Can also be axiomatized…
- But briefly

$$\boxed{\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)}}$$

$$\begin{aligned} \Pr(\text{Burglary}) &= .2 \\ \Pr(\text{Burglary}|\text{Earthquake}) &= .2 \end{aligned}$$

$$\begin{aligned} \Pr(\text{Alarm}) &= .2442 \\ \Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow \end{aligned}$$

[Darwiche '08]

Product Rule

Basic Properties of Probability

Betting Semantics

Inconsistent Beliefs

Agent 1		Agent 2		Outcomes and payoffs to Agent 1			
Proposition	Belief	Bet	Stakes	a, b	$a, \neg b$	$\neg a, b$	$\neg a, \neg b$
a	0.4	a	4 to 6	–6	–6	4	4
b	0.3	b	3 to 7	–7	3	–7	3
$a \vee b$	0.8	$\neg(a \vee b)$	2 to 8	2	2	2	–8
				–11	–1	–1	–1

Computing Probabilities: Example

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Monotonicity of Belief?

- Recall: monotonicity of logic
- Is it possible to observe something new and undo prior beliefs?

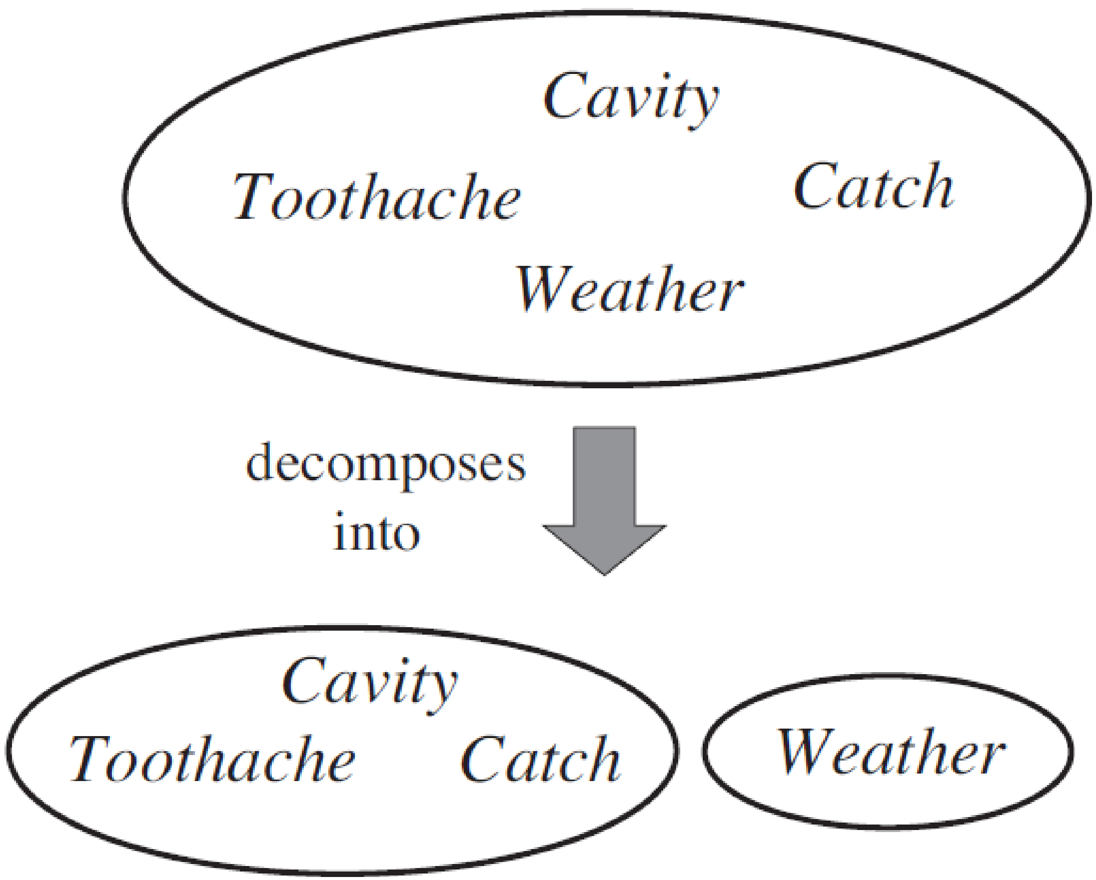
world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$\begin{aligned} \Pr(\text{Alarm}) &= .2442 \\ \Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow \end{aligned}$$

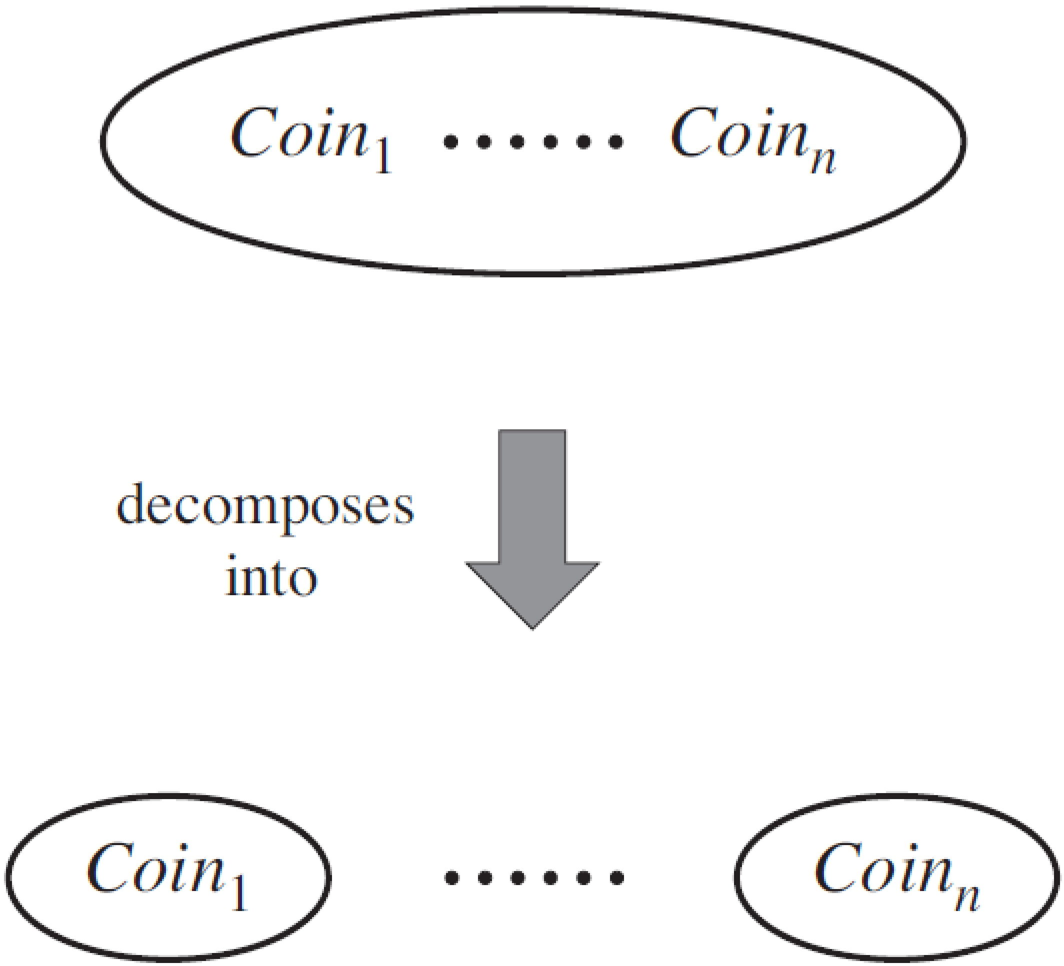
- Example:
  - Alarm and not Earthquake: .1620+.0072=0.1692
  - Not Earthquake: .9
  - Alarm given not Earthquake: .188

[Darwiche '08]

Independence



Independence

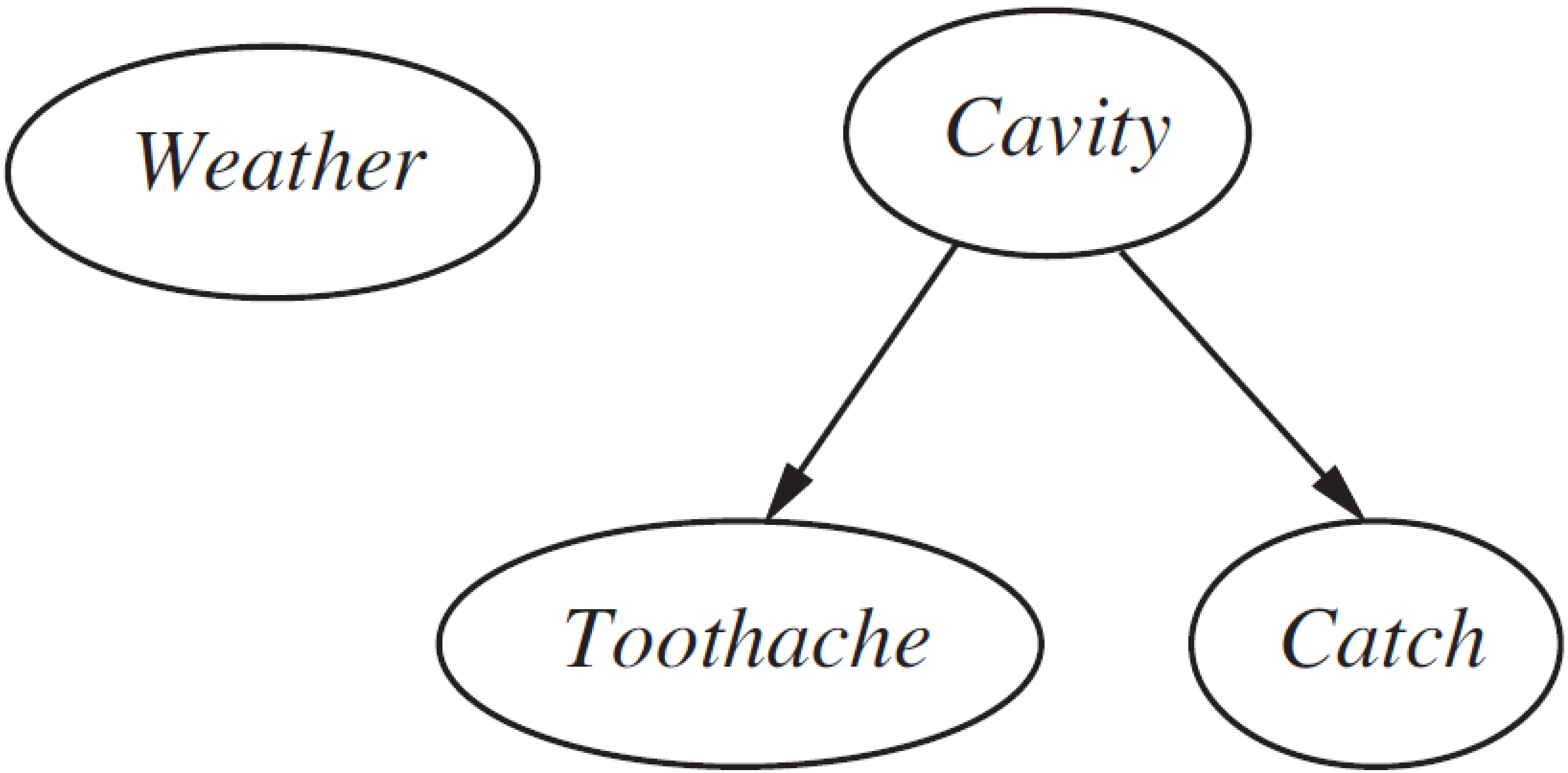


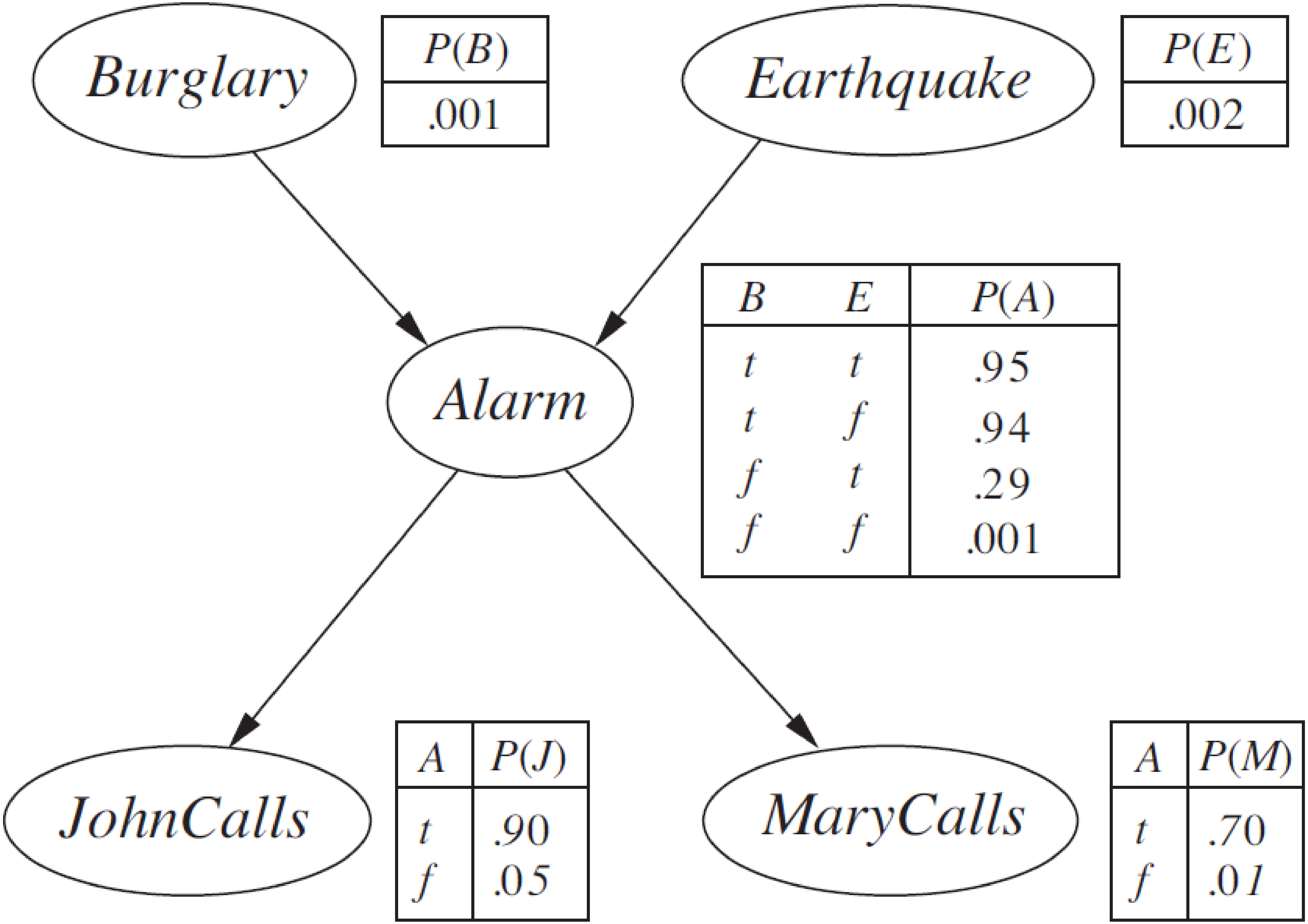
Naïve Bayes Assumption

$$\mathbf{P}(Cause, \mathit{Effect}_1, \ldots, \mathit{Effect}_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(\mathit{Effect}_i \mid Cause)$$

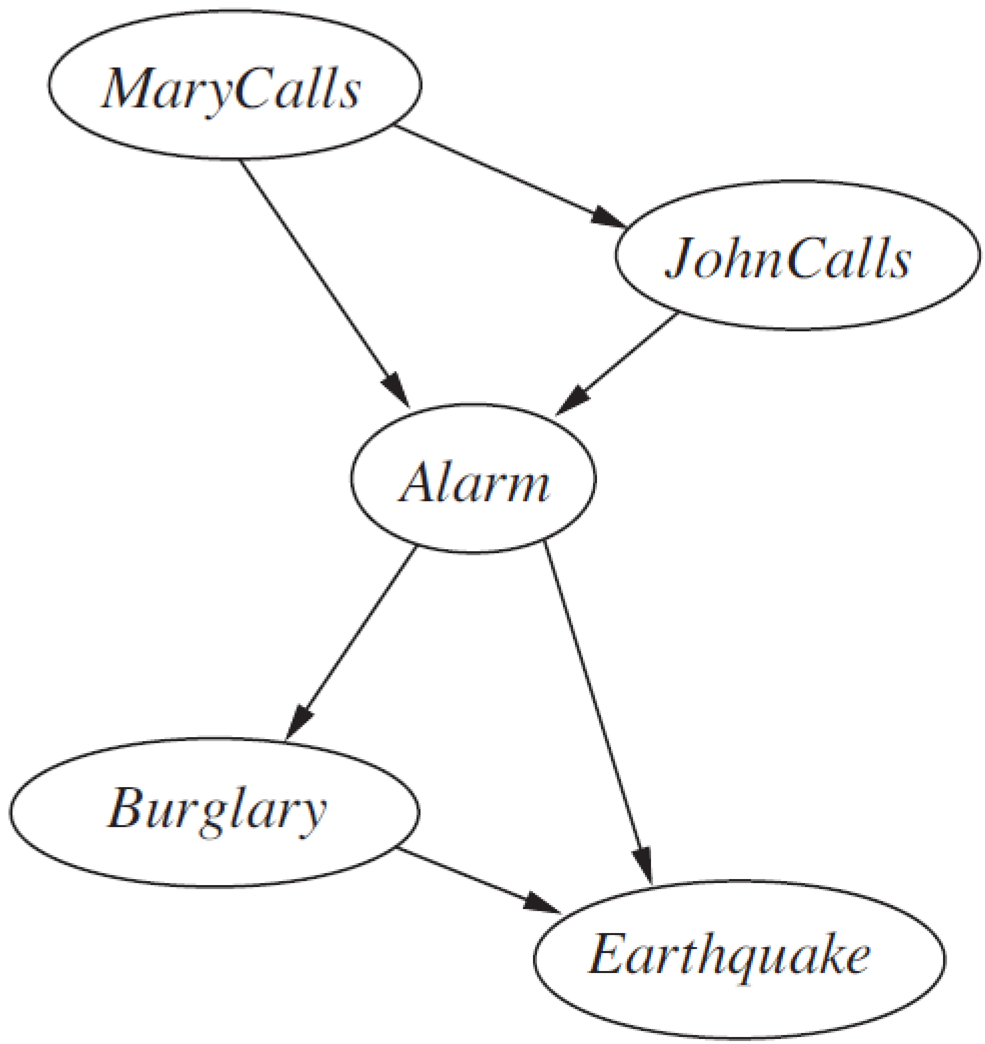
This is how spam filters work!

Bayesian Networks

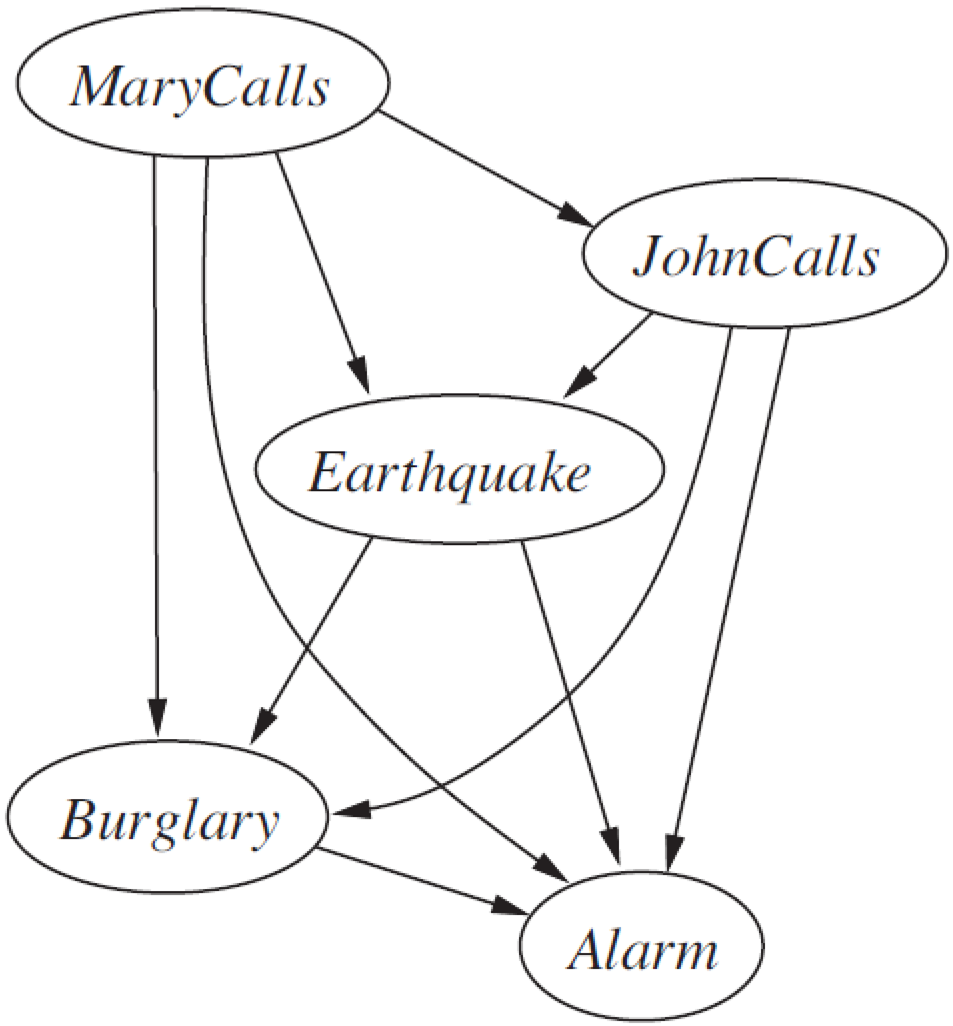



Burglary
P(B)
.001
Earthquake
P(E)
.002
Alarm
B E P(A)
t t .95
t f .94
f t .29
f f .001
JohnCalls
A P(J)
t .90
f .05
MaryCalls
A P(M)
t .70
f .01

Conditional Independence and Order



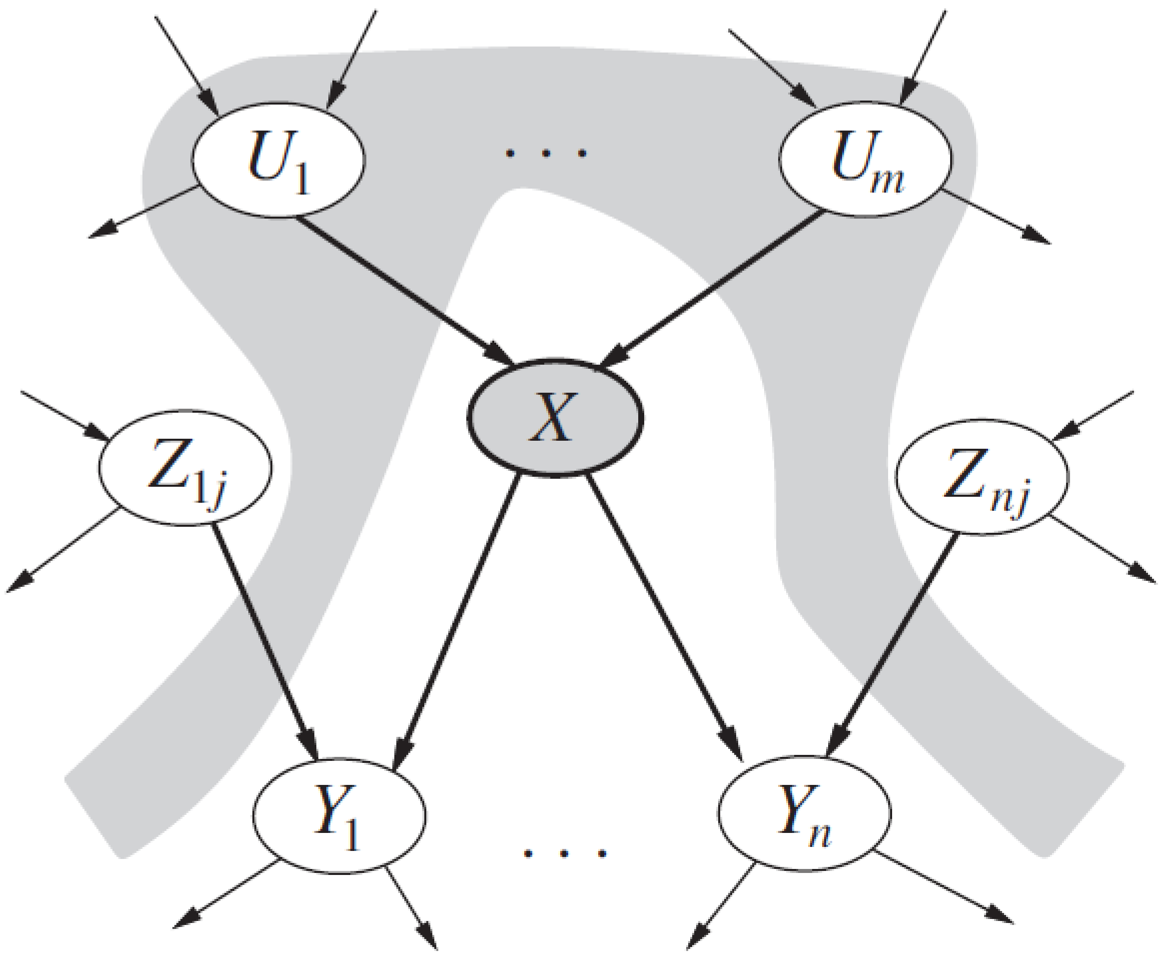
Conditional Independence and Order



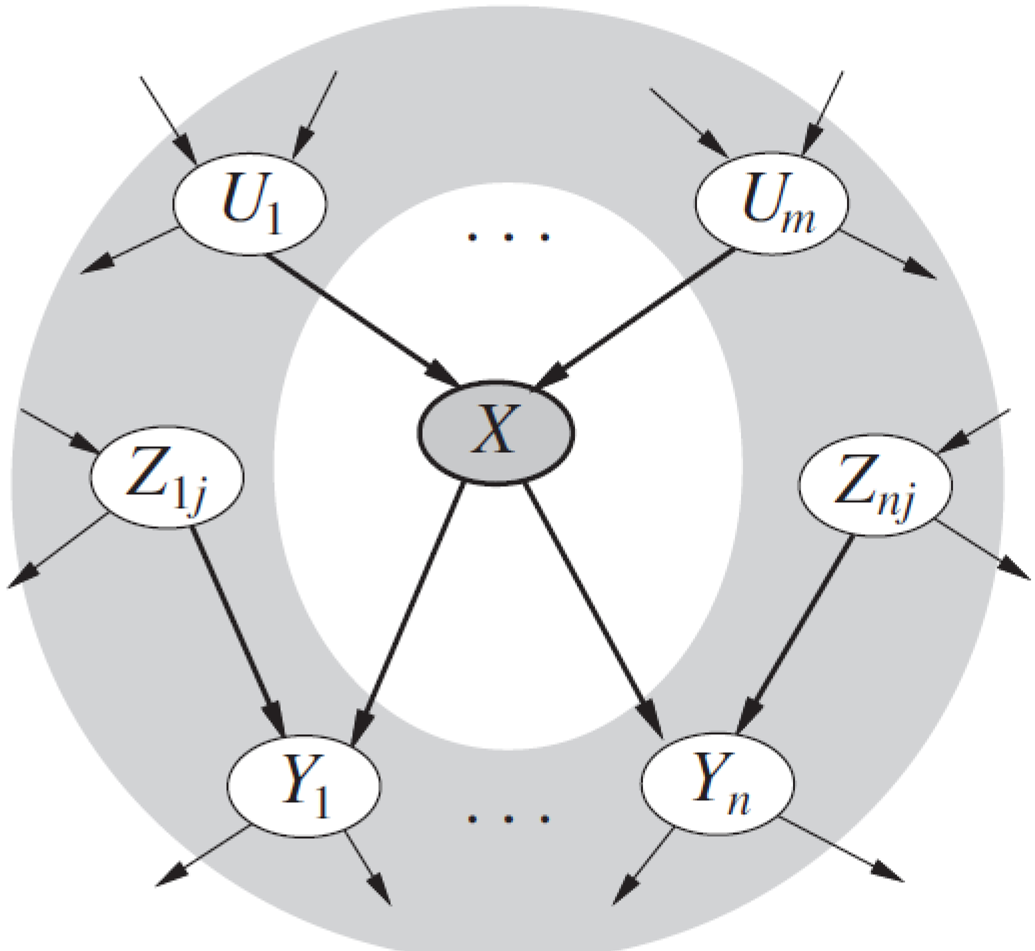
Topological Semantics

What knowledge is encoded in Bayesian network structure?

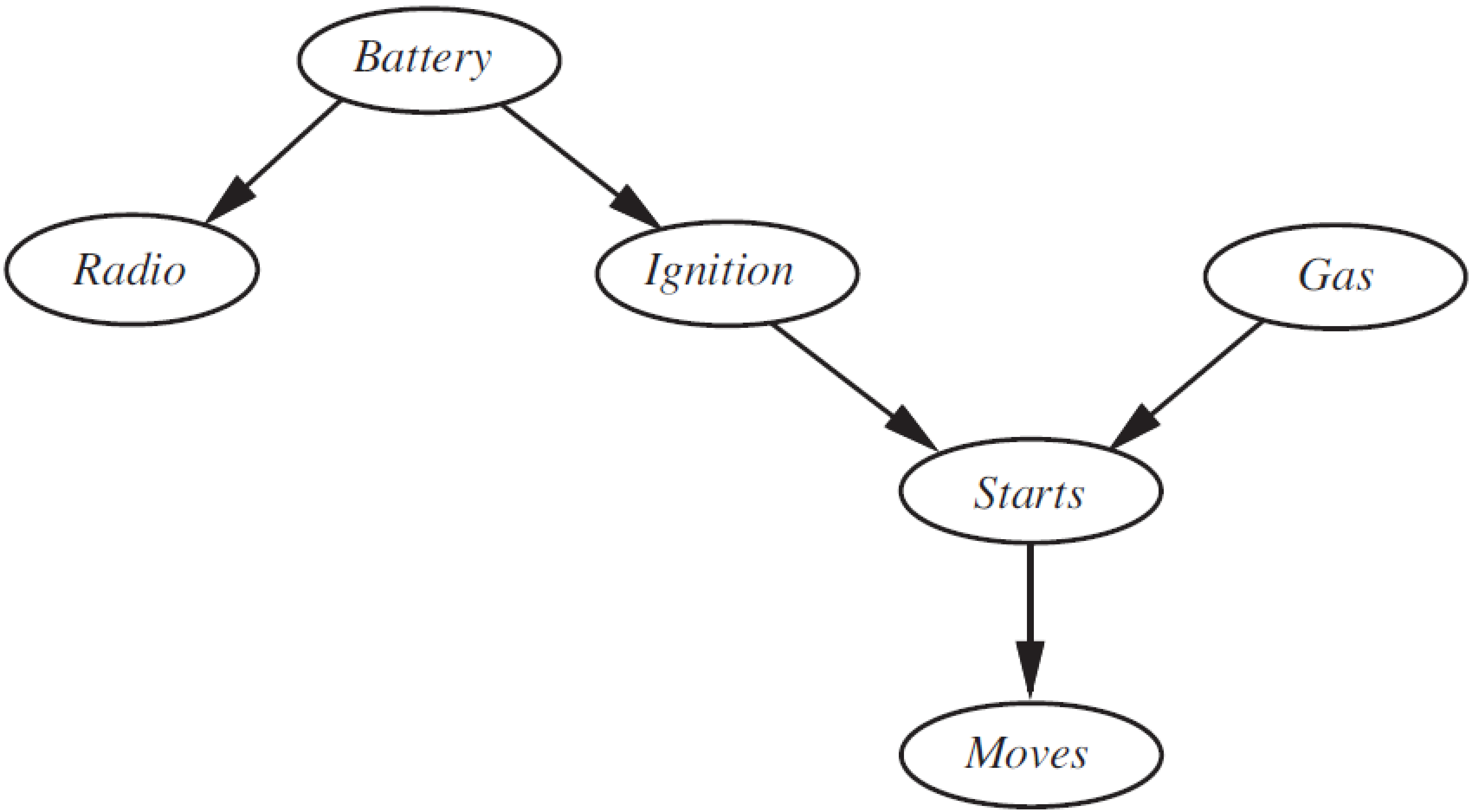
Markovian Assumptions



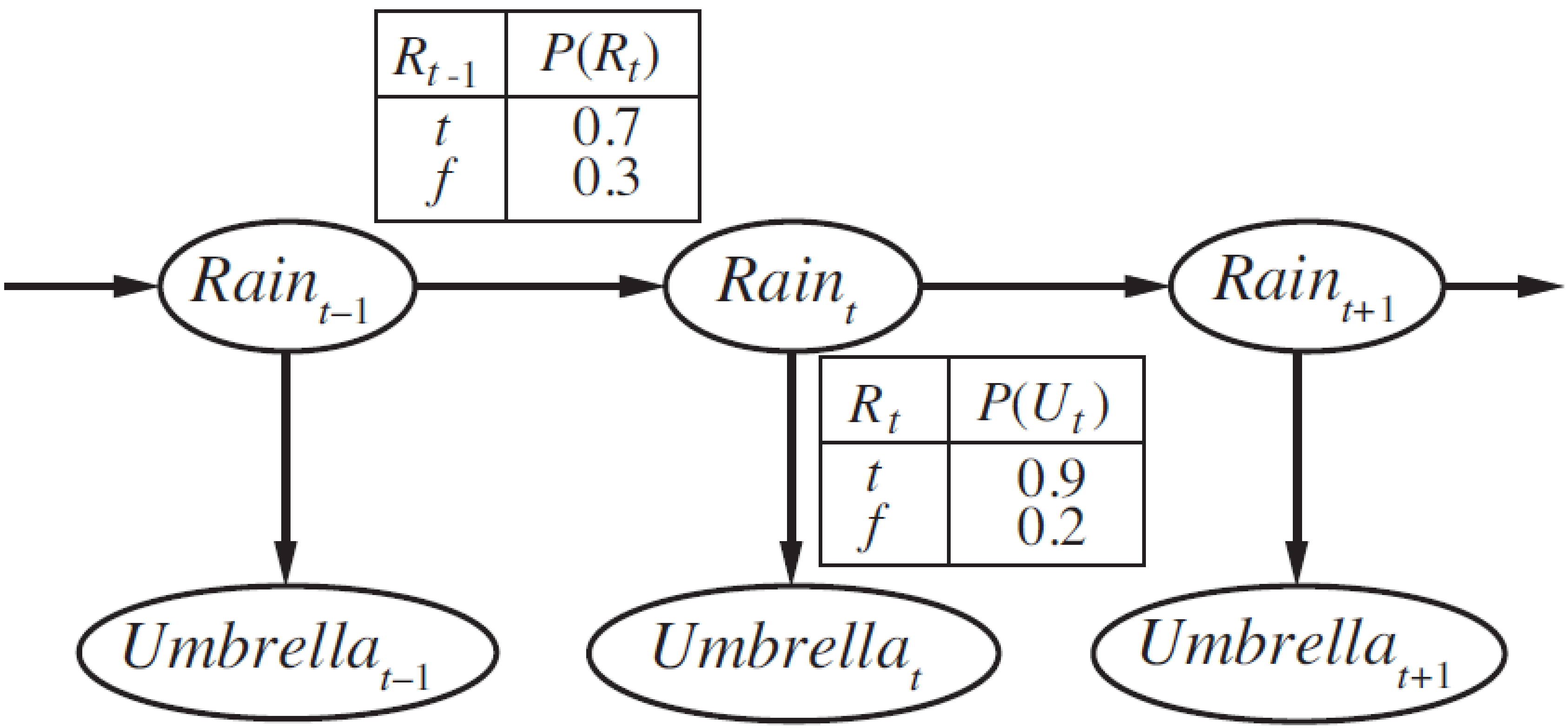
Markov Blanket



Example Network



Markov Chains
and Hidden Markov Models



Inference by Enumeration

Factors Multiplication

A	B	$\mathbf{f}_1(A,B)$
T	T	.3
T	F	.7
F	T	.9
F	F	.1

X

B	C	$\mathbf{f}_2(B,C)$
T	T	.2
T	F	.8
F	T	.6
F	F	.4

=

A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	T	$.3 \times .2 = .06$
T	T	F	$.3 \times .8 = .24$
T	F	T	$.7 \times .6 = .42$
T	F	F	$.7 \times .4 = .28$
F	T	T	$.9 \times .2 = .18$
F	T	F	$.9 \times .8 = .72$
F	F	T	$.1 \times .6 = .06$
F	F	F	$.1 \times .4 = .04$

Summing out Variable from Factor

A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	T	$.3 \times .2 = .06$
T	T	F	$.3 \times .8 = .24$
T	F	T	$.7 \times .6 = .42$
T	F	F	$.7 \times .4 = .28$
F	T	T	$.9 \times .2 = .18$
F	T	F	$.9 \times .8 = .72$
F	F	T	$.1 \times .6 = .06$
F	F	F	$.1 \times .4 = .04$

$$\begin{aligned}\mathbf{f}(B,C) &= \sum_a \mathbf{f}_3(A,B,C) = \mathbf{f}_3(a,B,C) + \mathbf{f}_3(\neg a,B,C) \\ &= \begin{pmatrix} .06 & .24 \\ .42 & .28 \end{pmatrix} + \begin{pmatrix} .18 & .72 \\ .06 & .04 \end{pmatrix} = \begin{pmatrix} .24 & .96 \\ .48 & .32 \end{pmatrix}.\end{aligned}$$

Variable Elimination