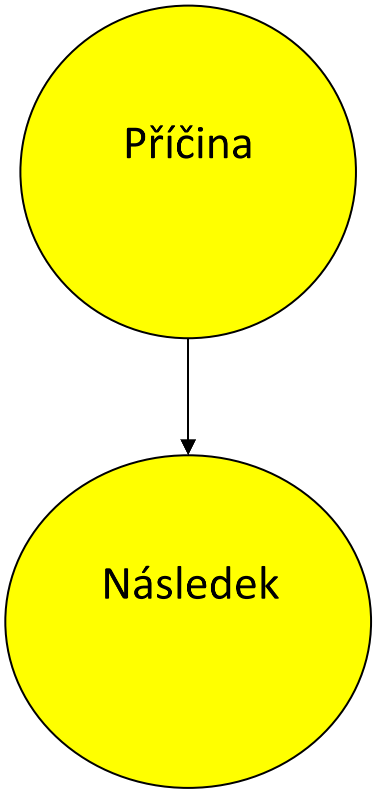
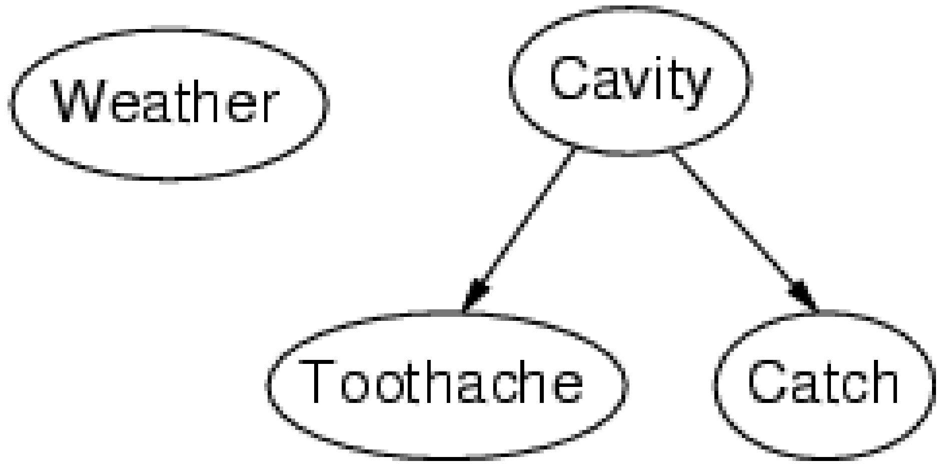
Bayesian networks



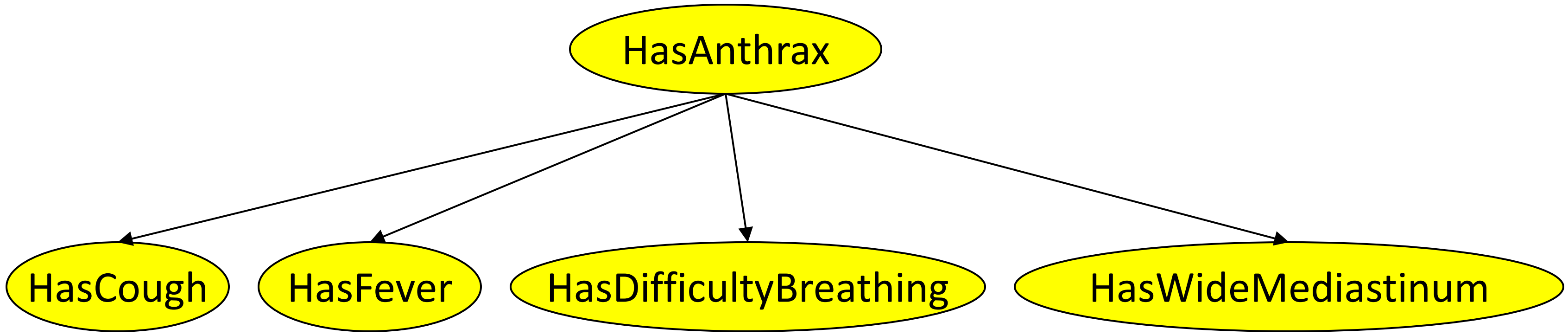
Example

- Topology of network encodes conditional independence assertions:



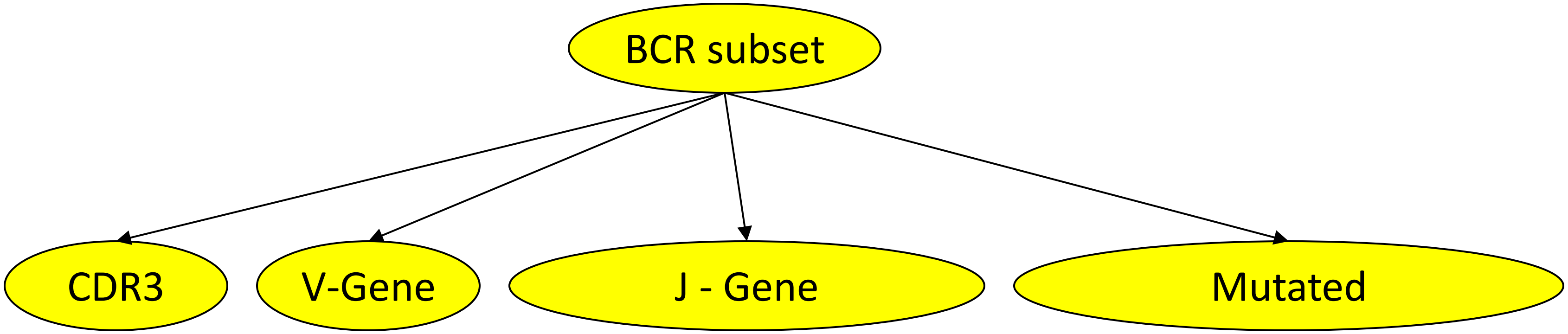
- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

Bayesian Networks



- In the opinion of many AI researchers, Bayesian networks are the most significant contribution in AI in the last 10 years
- They are used in many applications eg. spam filtering, speech recognition, robotics, diagnostic systems and even syndromic surveillance

Example



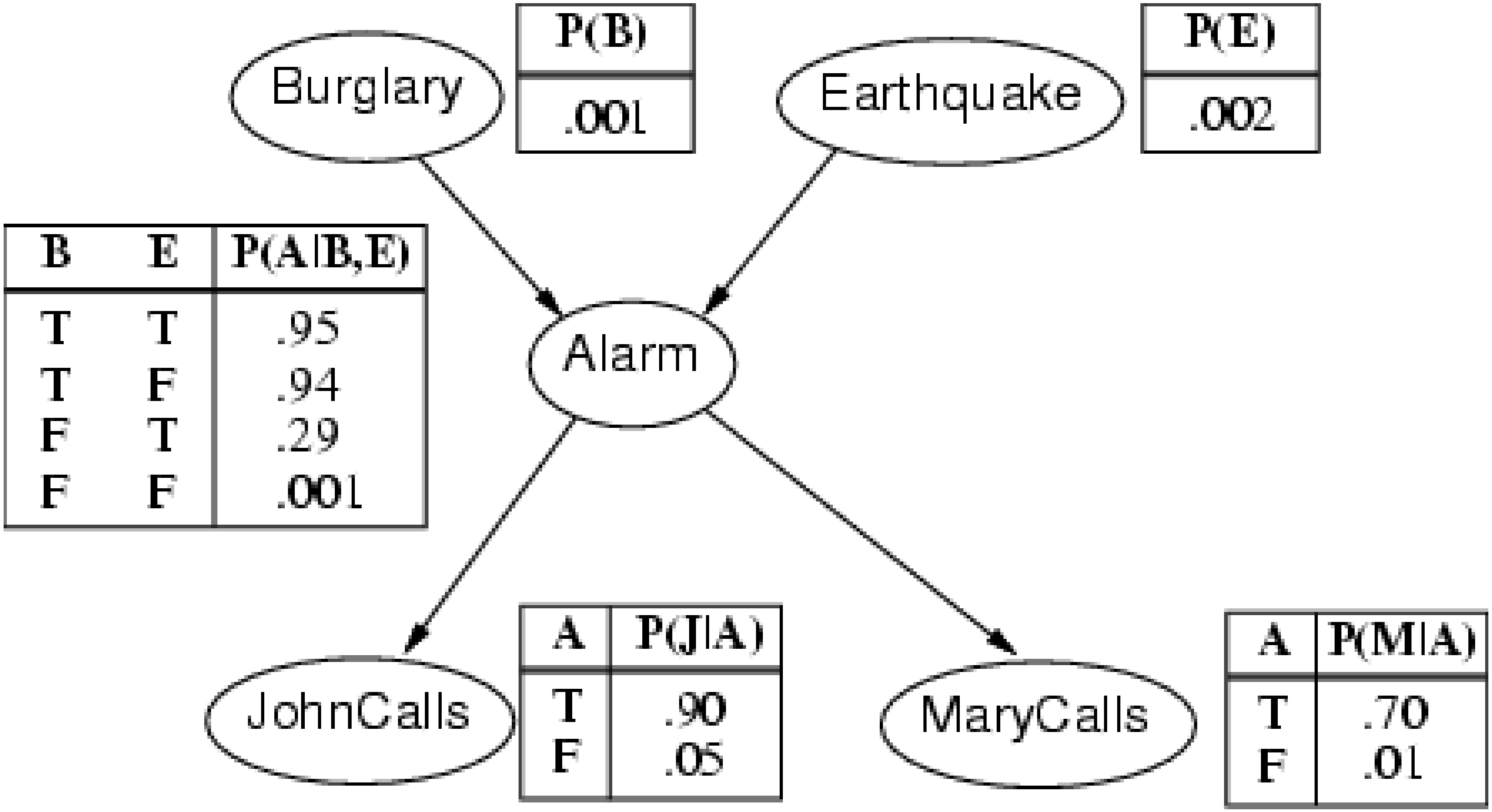
Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- Network topology reflects "causal" knowledge:
  - 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 contd.



P(B)
.001

P(E)
.002

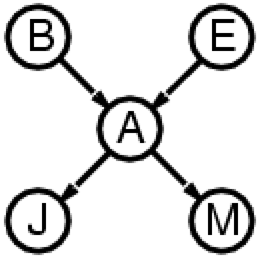
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

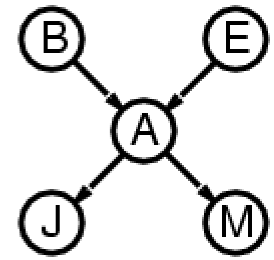


- Each row requires one number p for $X_i = true$ (the number for $X_i = false$ is just $1-p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

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



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$$

Constructing Bayesian networks

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

$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

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

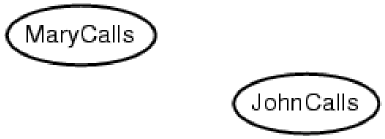
(chain rule)

$$= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

(by construction)

Example

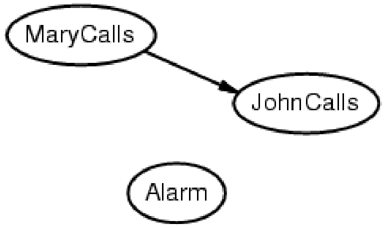
- Suppose we choose the ordering *M, J, A, B, E*
-



P$(J \mid M) = $***P***$(J)$?

Example

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



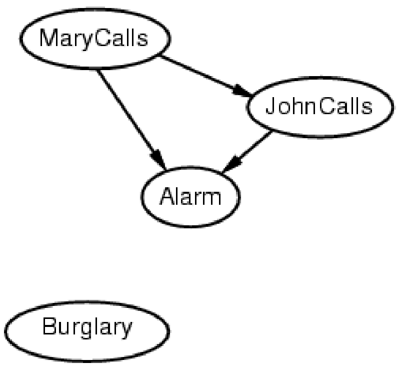
P*(J | M) =* ***P****(J)?*

No

P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*?

Example

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



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

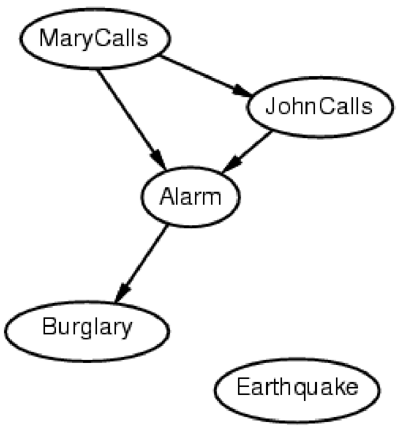
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$?

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$?

Example

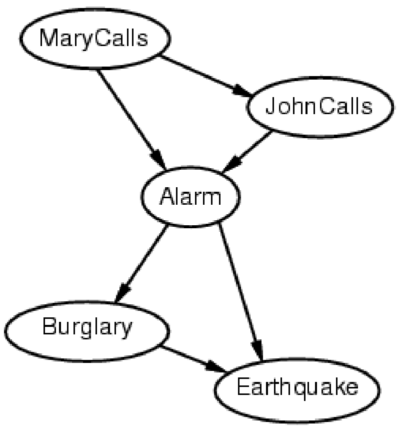
- Suppose we choose the ordering M, J, A, B, E
-



P(*J* | *M*) = ***P***(*J*)?
No
P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**
P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)? **Yes**
P(*B* | *A*, *J*, *M*) = ***P***(*B*)? **No**
P(*E* | *B*, *A* ,*J*, *M*) = ***P***(*E* | *A*)?
P(*E* | *B*, *A*, *J*, *M*) = ***P***(*E* | *A*, *B*)?

Example

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



P*(J | M) =* ***P****(J)?*

No

P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*? **No**

P*(B | A, J, M) =* ***P****(B | A)*? **Yes**

P*(B | A, J, M) =* ***P****(B)*? **No**

P*(E | B, A ,J, M) =* ***P****(E | A)*? **No**

P*(E | B, A, J, M) =* ***P****(E | A, B)*? **Yes**

Outline

1. Introduction
2. Probability Primer
3. Bayesian networks
4. Bayesian networks in syndromic surveillance

Probability Primer: Random Variables

- A random variable is the basic element of probability
- Refers to an event and there is some degree of uncertainty as to the outcome of the event
- For example, the random variable *A* could be the event of getting a heads on a coin flip

Boolean Random Variables

- We will start with the simplest type of random variables – Boolean ones
- Take the values *true* or *false*
- Think of the event as occurring or not occurring
- Examples (Let A be a Boolean random variable):

A = Getting heads on a coin flip

A = It will rain today

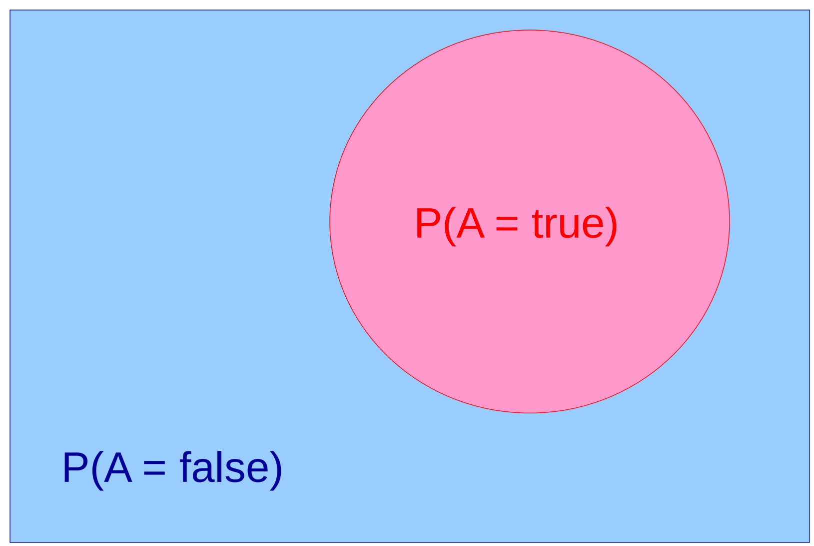
A = The Cubs win the World Series in 2007

Probabilities

We will write *P(A = true)* to mean the probability that *A = true*.

What is probability? It is the relative frequency with which an outcome would be obtained if the process were repeated a large number of times under similar conditions*

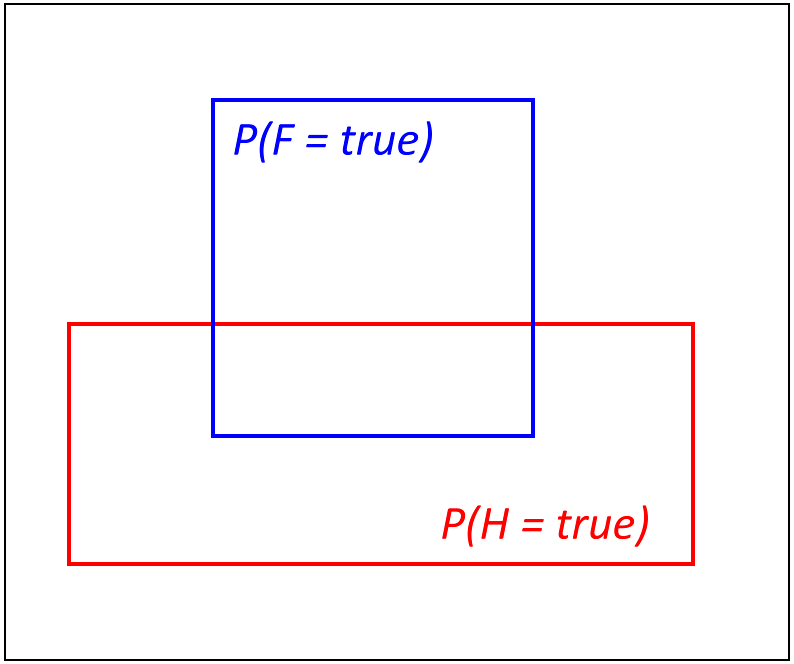
The sum of the red and blue areas is 1



*Ahem…there's also the Bayesian definition which says probability is your degree of belief in an outcome

Conditional Probability

- P(*A = true* | *B = true*) = Out of all the outcomes in which *B* is true, how many also have *A* equal to true
- Read this as: "Probability of *A* conditioned on *B*" or "Probability of *A* given *B*"



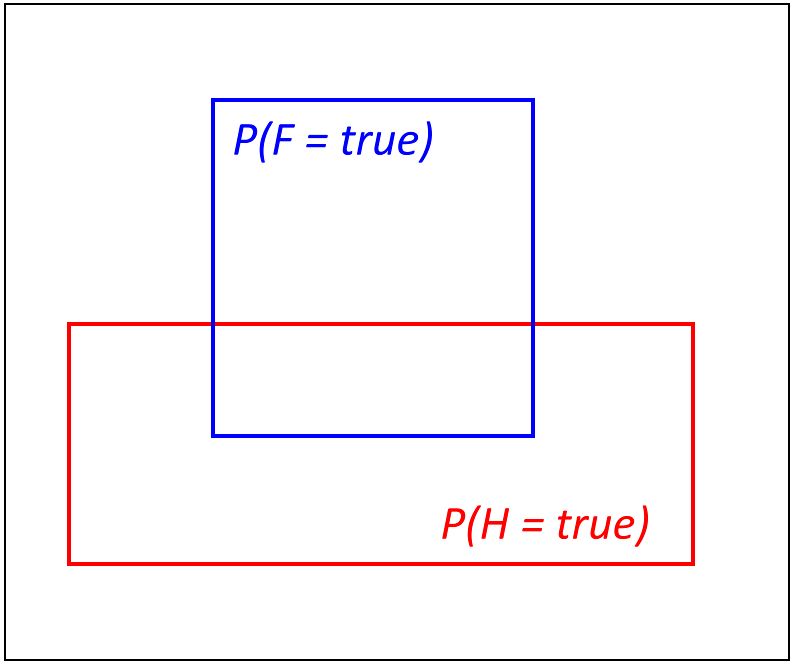
H = "Have a headache"
F = "Coming down with Flu"

P(*H = true*) = 1/10
P(*F = true*) = 1/40
P(*H = true* | *F = true*) = 1/2

"Headaches are rare and flu is rarer, but if you're coming down with flu there's a 50-50 chance you'll have a headache."

The Joint Probability Distribution

- We will write *P(A = true, B = true)* to mean "the probability of *A = true* **and** *B = true*"
- Notice that:



$$P(H{=}true \mid F{=}true)$$

$$= \frac{\text{Area of "H and F" region}}{\text{Area of "F" region}}$$

$$= \frac{P(H = true, F = true)}{P(F = true)}$$

In general, $P(X|Y)=P(X,Y)/P(Y)$

The Joint Probability Distribution

- Joint probabilities can be between any number of variables

  eg. *P(A = true, B = true, C = true)*
- For each combination of variables, we need to say how probable that combination is
- The probabilities of these combinations need to sum to 1

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Sums to 1

The Joint Probability Distribution

- Once you have the joint probability distribution, you can calculate any probability involving *A*, *B*, and *C*
- Note: May need to use marginalization and Bayes rule, (both of which are not discussed in these slides)

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Examples of things you can compute:

- $P(A=true)$ = sum of $P(A,B,C)$ in rows with $A=true$
- $P(A=true, B = true \mid C=true) =$

  $P(A = true, B = true, C = true) / P(C = true)$

The Problem with the Joint Distribution

- Lots of entries in the table to fill up!
- For *k* Boolean random variables, you need a table of size 2^k
- How do we use fewer numbers? Need the concept of independence

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Independence

Variables *A* and *B* are independent if any of the following hold:

- $P(A,B) = P(A)\,P(B)$
- $P(A \mid B) = P(A)$
- $P(B \mid A) = P(B)$

This says that knowing the outcome of *A* does not tell me anything new about the outcome of *B*.

Independence

How is independence useful?

- Suppose you have n coin flips and you want to calculate the joint distribution $\boldsymbol{P}(C_1, \ldots, C_n)$
- If the coin flips are not independent, you need 2^n values in the table
- If the coin flips are independent, then

$$P(C_1, \ldots, C_n) = \prod_{i=1}^{n} P(C_i)$$

Each $P(C_i)$ table has 2 entries and there are n of them for a total of $2n$ values

Conditional Independence

Variables *A* and *B* are conditionally independent given *C* if any of the following hold:

- $P(A, B \mid C) = P(A \mid C)\, P(B \mid C)$
- $P(A \mid B, C) = P(A \mid C)$
- $P(B \mid A, C) = P(B \mid C)$

Knowing *C* tells me everything about *B*. I don't gain anything by knowing *A* (either because *A* doesn't influence *B* or because knowing *C* provides all the information knowing *A* would give)

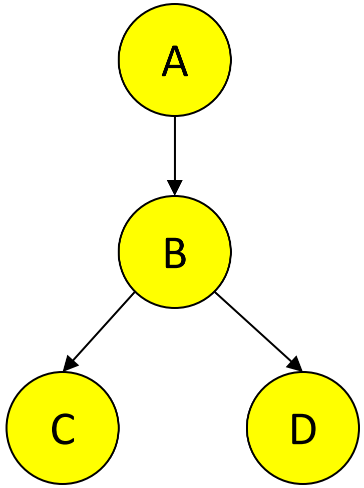
Outline

1. Introduction
2. Probability Primer
3. Bayesian networks
4. Bayesian networks in syndromic surveillance

A Bayesian Network

A Bayesian network is made up of:

1. A Directed Acyclic Graph



2. A set of tables for each node in the graph

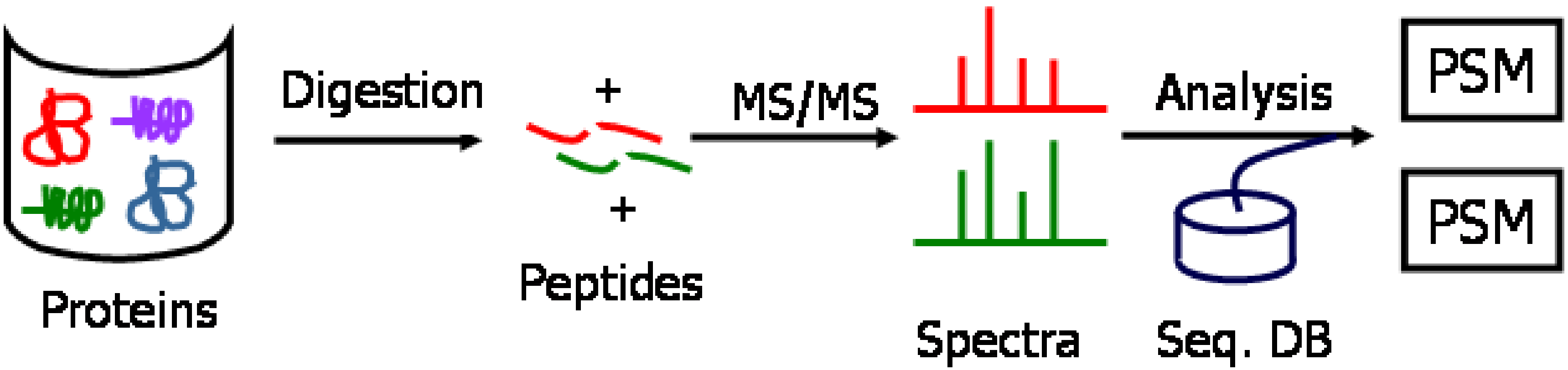
A	P(A)
false	0.6
true	0.4

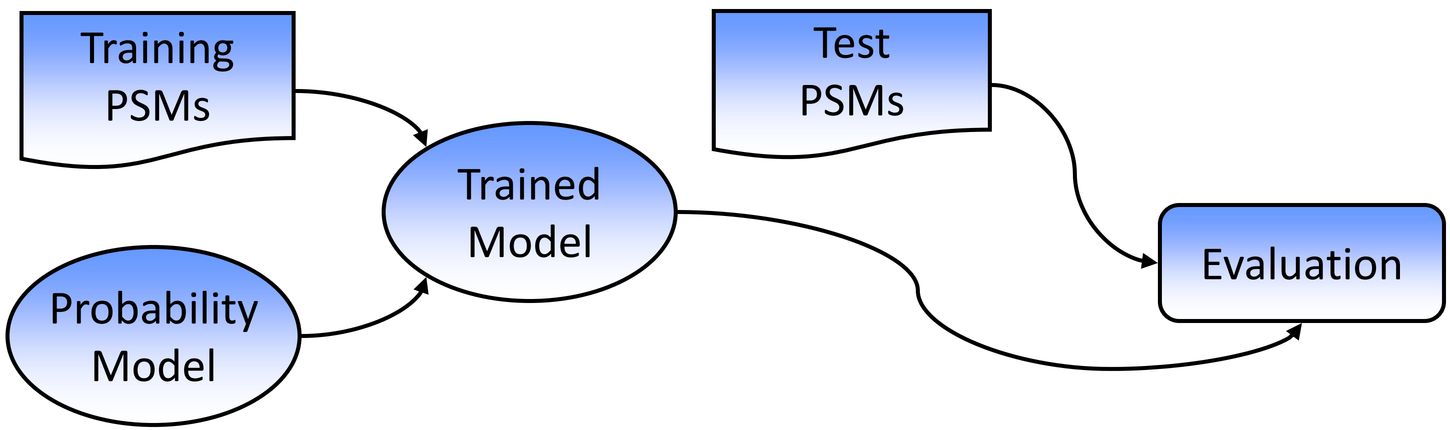
A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

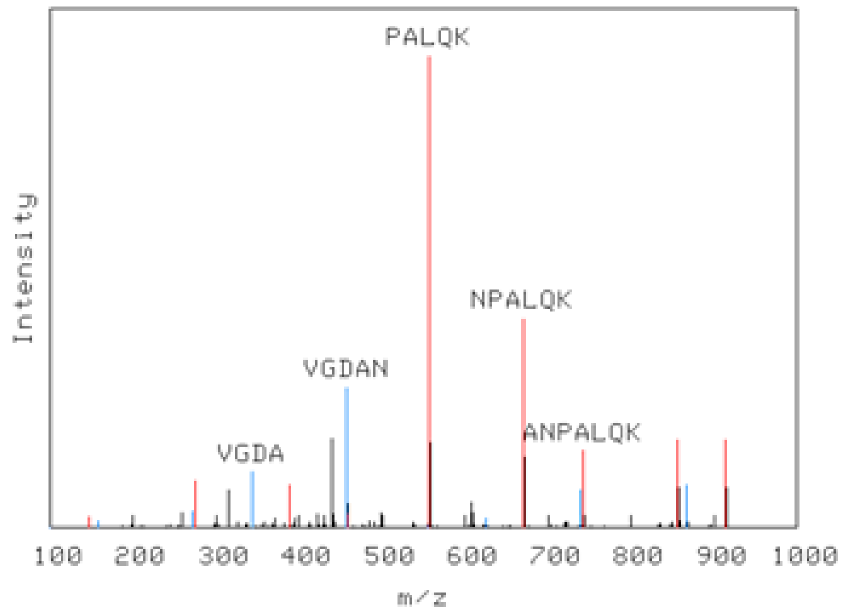
Shotgun proteomics

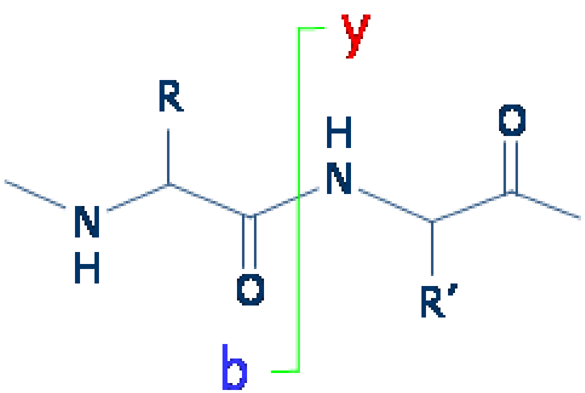




PSM = peptide-spectrum match

Peptide sequence influences peak height





VGDANPALQK

V + GDANPALQK

VG + DANPALQK

VGD + ANPALQK

VGDA + NPALQK

VGDAN + PALQK

VGDANP + ALQK

VGDANPA + LQK

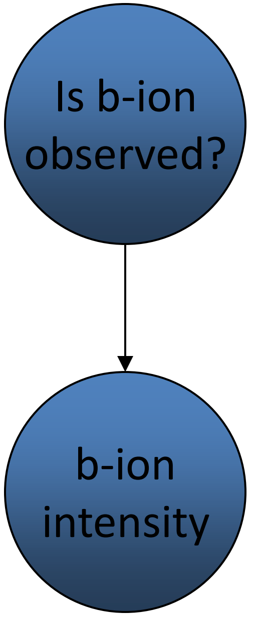
VGDANPAL + QK

VGDANPALQ + K

b-ions y-ions

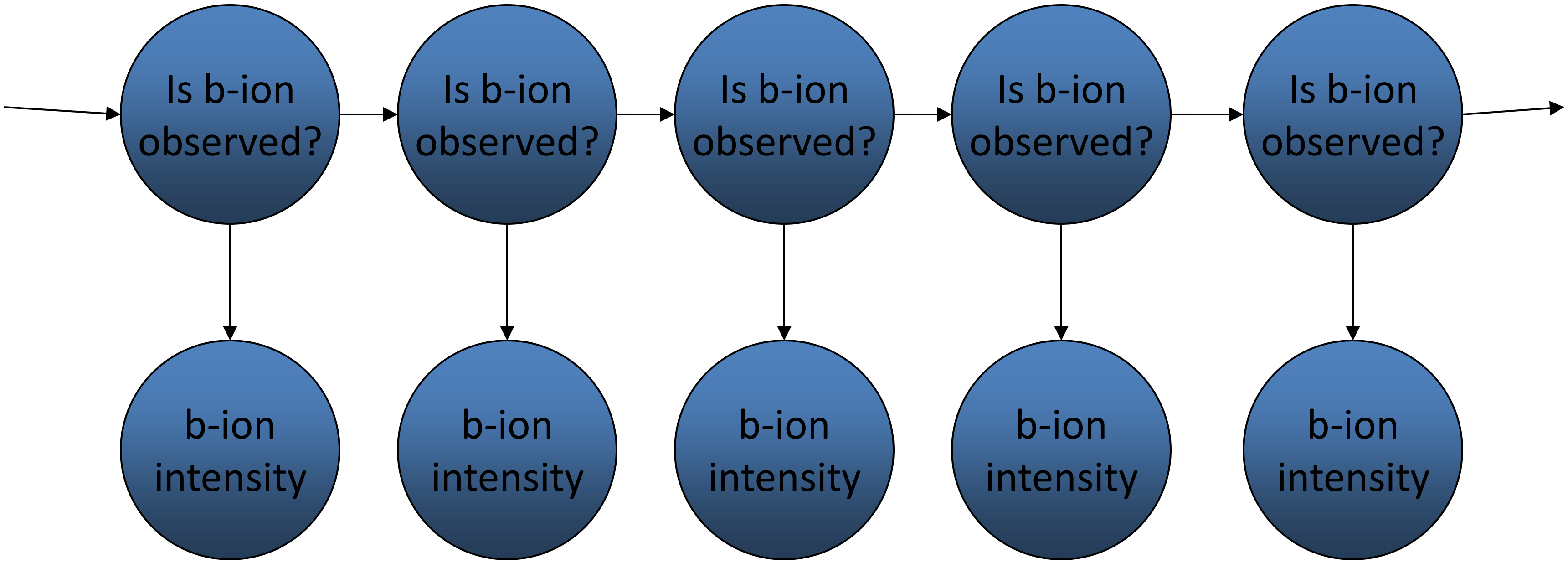
Bayesian network

- We model peptide fragmentation using a Bayesian network.
- Nodes represent random variables, and edges represent conditional dependencies.
- Each node stores a conditional probability table (CPT) giving Pr(node|parents).



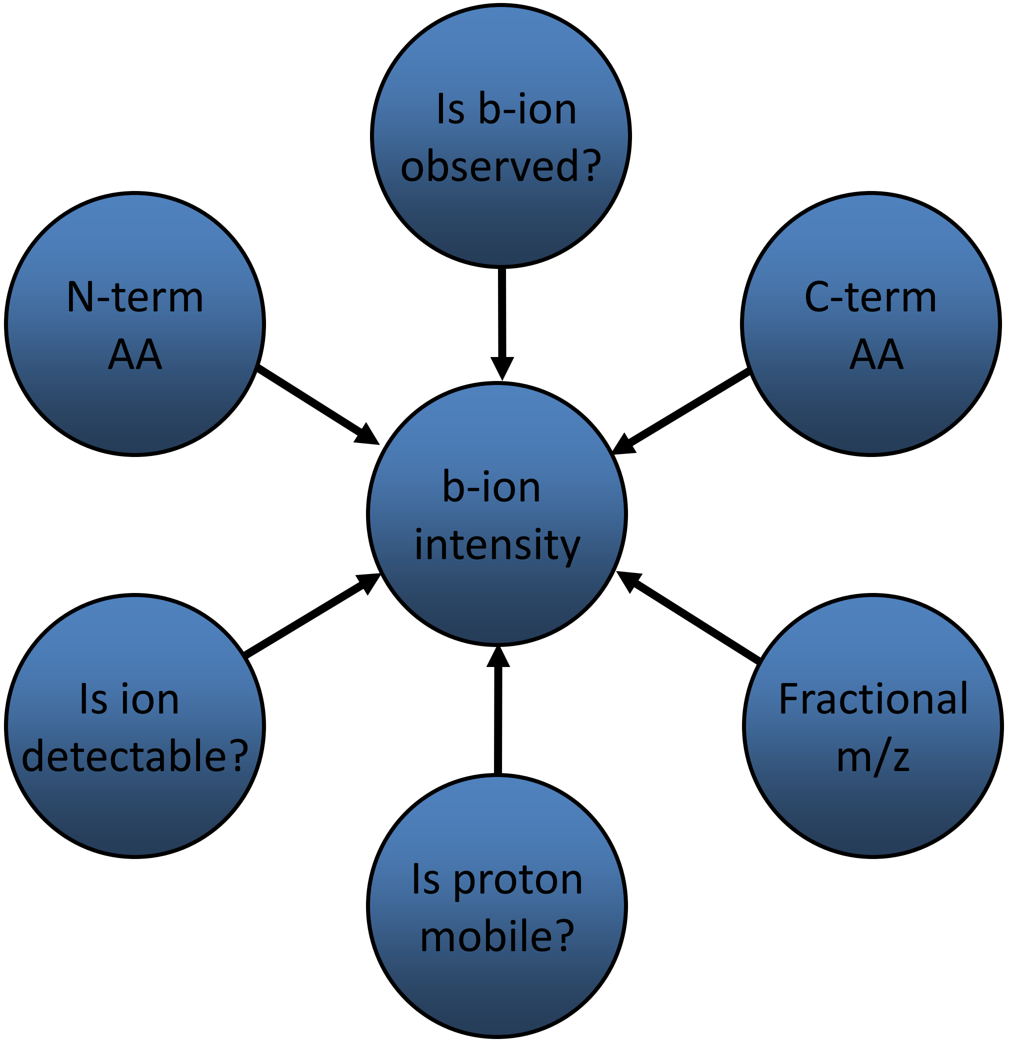
	intensity > 50%	intensity < 50%
b-ion observed	0.25	0.75
no b-ion observed	0.00	1.00

Ion series modeled in a Markov chain

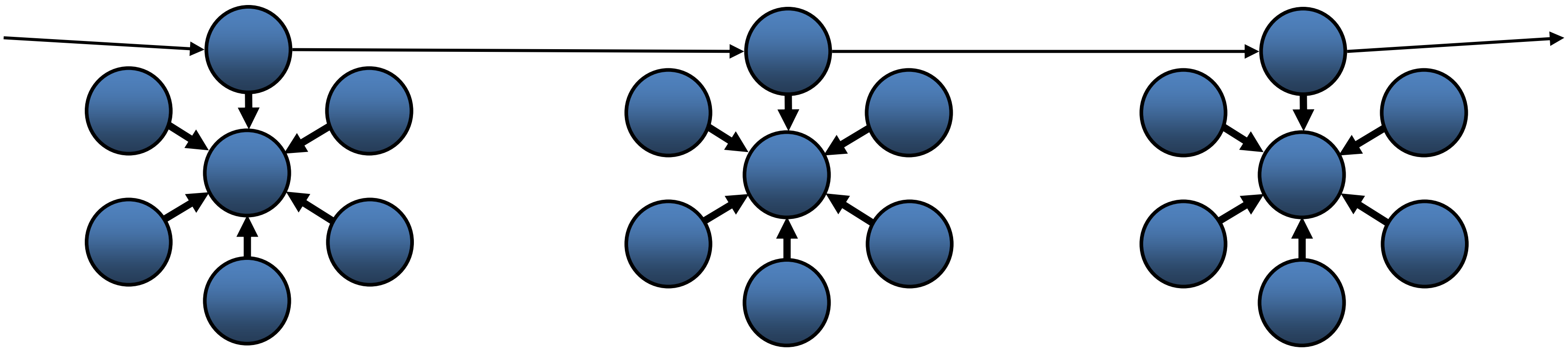


~ PepHMM (Han et al., 2005).

A more realistic model



Ion series modeled in a Markov chain



$$LOR_b = \log \frac{\Pr(\text{b-ions, peptide} \mid \text{model})}{\Pr(\text{b-ions, peptide} \mid \text{null model})}$$