

Reasoning with Graphical Models

Slides Set 3b:

# Building Bayesian Networks

*Rina Dechter*

Darwiche chapters 5 (a sneak preview)

slides3b 276 2024

# Outline

DAGS, Markov(G), Bayesian networks

Graphoids: axioms of for inferring conditional independence (CI)

D-separation: Inferring CIs in graphs

I-maps, D-maps, perfect maps

**Construction a minimal I-map of a distribution**

Markov boundary and blanket

Markov networks

# Bayesian Networks as Knowledge-Bases

Given any distribution,  $P$ , and an ordering we can construct a minimal i-map.

The conditional probabilities of  $x$  given its parents is all we need.

In practice we go in the opposite direction: the parents must be identified by human expert... they can be viewed as direct causes, or direct influences.

## BAYESIAN NETWORK AS A KNOWLEDGE BASE

### STRUCTURING THE NETWORK

- Given any joint distribution  $P(x_1, \dots, x_n)$  and an ordering  $d$  of the variables in  $U$ , Corollary 4 prescribes a simple recursive procedure for constructing a Bayesian network.
- Choose  $X_1$  as a root and assign to it the marginal probability  $P(x_1)$  dictated by  $P(x_1, \dots, x_n)$ .
- If  $X_2$  is dependent on  $X_1$ , a link from  $X_1$  to  $X_2$  is established and quantified by  $P(x_2|x_1)$ . Otherwise, we leave  $X_1$  and  $X_2$  unconnected and assign the prior probability  $P(x_2)$  to node  $X_2$ .
- At the  $i$ -th stage, we form the node  $X_i$ , draw a group of directed links to  $X_i$  from a parent set  $\Pi_{X_i}$  defined by Eq. (3.27), and quantify this group of links by the conditional probability  $P(x_i | \mathbf{n}_{X_i})$ .
- The result is a directed acyclic graph that represents all the independencies that follow from the definitions of the parent sets.

- In practice,  $P(x_1, \dots, x_n)$  is not available.
- The parent sets  $\Pi_{X_i}$  must be identified by human judgment.
- To specify the strengths of influences, assess the conditional probabilities  $P(x_i | \Pi_{X_i})$  by some functions  $F_i(x_i, \Pi_{X_i})$  and make sure these assessments satisfy

$$\sum_{x_i} F_i(x_i, \Pi_{X_i}) = 1, \quad (3.30)$$

where  $0 \leq F_i(x_i, \Pi_{X_i}) \leq 1$

- This specification is complete and consistent because the product form

$$P_a(x_1, \dots, x_n) = \prod_i F_i(x_i, \Pi_{X_i}) \quad (3.31)$$

constitutes a joint probability distribution that supports the assessed quantities.

$$P_a(x_i | \Pi_{X_i}) = \frac{P_a(x_i, \Pi_{X_i})}{P_a(\Pi_{X_i})} = \frac{\sum_{x_j \in (x_i \cup \Pi_{X_i})} P_a(x_1, \dots, x_n)}{\sum_{x_j \in \Pi_{X_i}} P_a(x_1, \dots, x_n)} = F_i(x_i, \Pi_{X_i}) \quad (3.32)$$

- DAGs constructed by this method will be called *Bayesian belief networks* or *causal networks* interchangeably.

# Outline

DAGS, Markov(G), Bayesian networks

Graphoids: axioms of for inferring conditional independence (CI)

**D-separation: Inferring CIs in graphs**

Soundness, completeness of d-seperation

I-maps, D-maps, perfect maps

Construction a minimal I-map of a distribution

**Markov boundary and blanket**

Markov networks

# Blankets and Boundaries

## Definition

Let  $P_{\mathbf{r}}$  be a distribution over variables  $\mathbf{X}$ . A **Markov blanket** for a variable  $X \in \mathbf{X}$  is a set of variables  $\mathbf{B} \subseteq \mathbf{X}$  such that  $X \notin \mathbf{B}$  and  $I_{P_{\mathbf{r}}}(X, \mathbf{B}, \mathbf{X} \setminus \mathbf{B} \setminus \{X\})$ .

A Markov blanket for  $X$  is a set of variables which, when known, will render every other variable irrelevant to  $X$ .

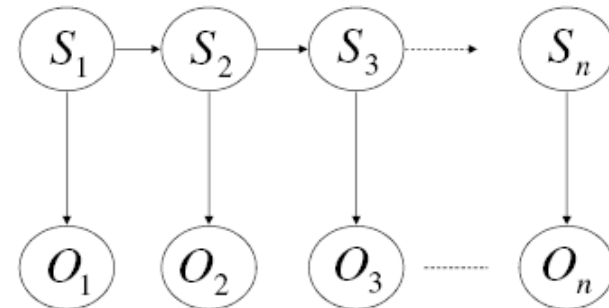
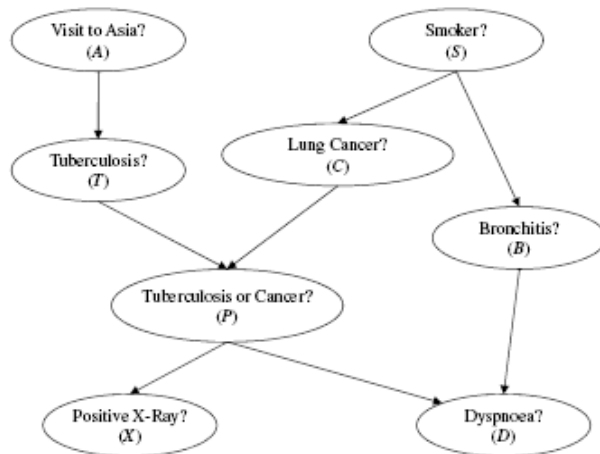
## Definition

A Markov blanket  $\mathbf{B}$  is **minimal** iff no strict subset of  $\mathbf{B}$  is also a Markov blanket. A minimal Markov blanket is a **Markov Boundary**.

The Markov Boundary for a variable is not unique, unless the distribution is strictly positive.

## Blanket Examples

If  $\Pr$  is induced by DAG  $G$ , then a Markov blanket for variable  $X$  with respect to  $\Pr$  can be constructed using its parents, children, and spouses in DAG  $G$ . Here, variable  $Y$  is a spouse of  $X$  if the two variables have a common child in DAG  $G$ .



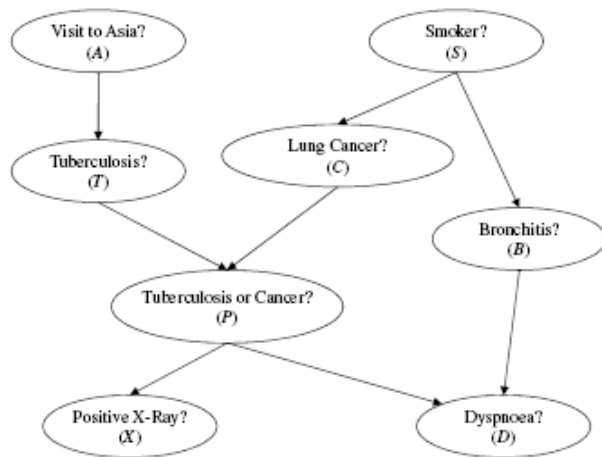
$\{S_{t-1}, S_{t+1}, O_t\}$  is a Markov blanket for every variable  $S_t$ , where  $t > 1$

What is a Markov blanket of C?

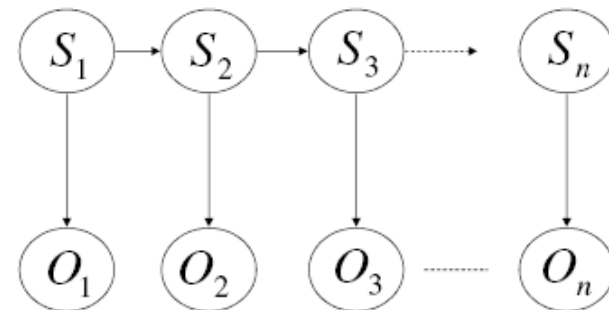


## Blanket Examples

If  $\Pr$  is induced by DAG  $G$ , then a Markov blanket for variable  $X$  with respect to  $\Pr$  can be constructed using its parents, children, and spouses in DAG  $G$ . Here, variable  $Y$  is a spouse of  $X$  if the two variables have a common child in DAG  $G$ .

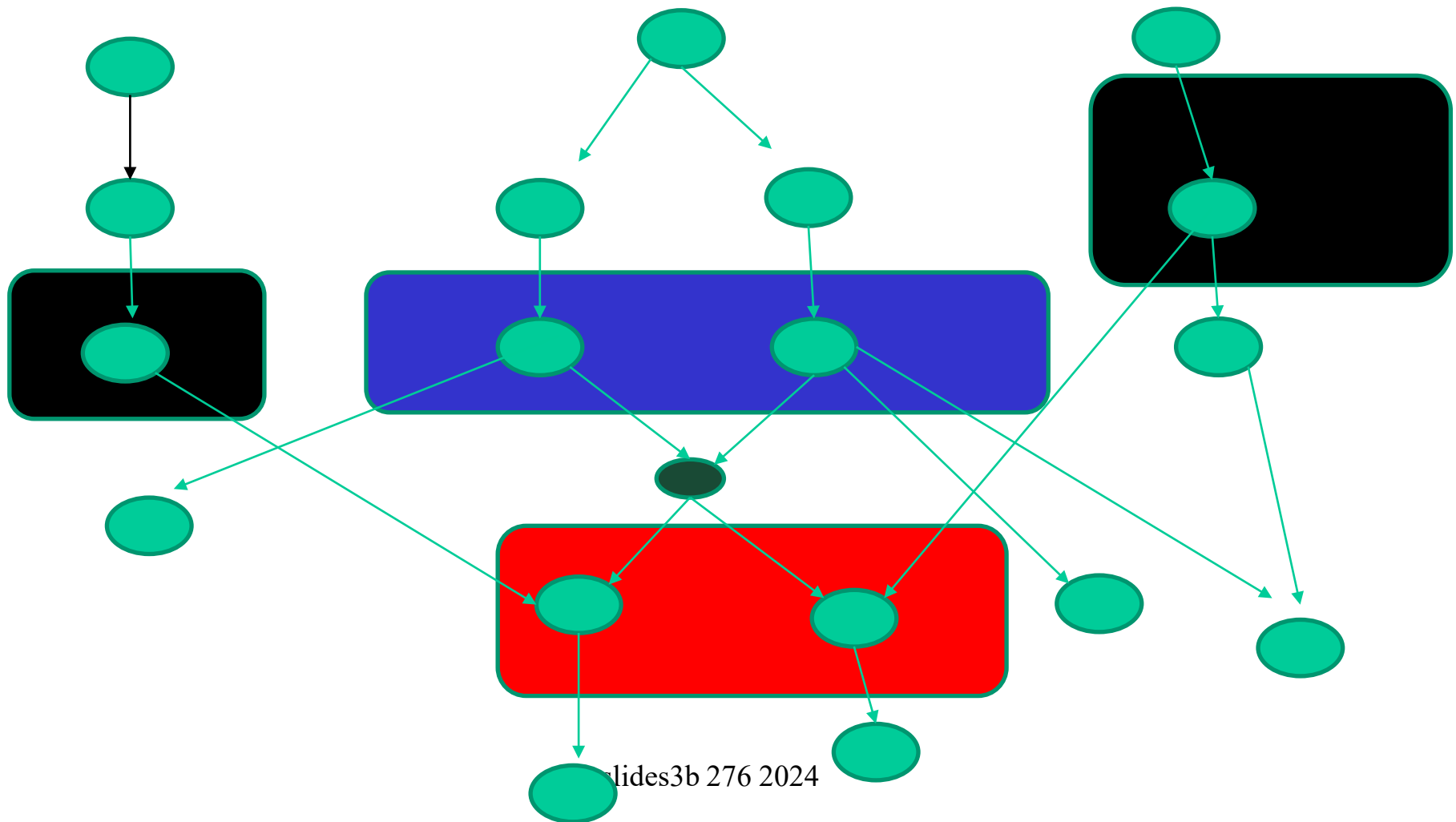


$\{S, P, T\}$  is a Markov blanket for variable  $C$



$\{S_{t-1}, S_{t+1}, O_t\}$  is a Markov blanket for every variable  $S_t$ , where  $t > 1$

# Markov Blanket



# Outline

DAGS, Markov(G), Bayesian networks

Graphoids: axioms of for inferring conditional independence (CI)

**D-separation: Inferring CIs in graphs**

Soundness, completeness of d-seperation

I-maps, D-maps, perfect maps

Construction a minimal I-map of a distribution

Markov boundary and blanket

**Markov networks, Markov Random Fields**

# Undirected Graphs as I-maps of Distributions

- We say  $\langle X, Z, Y \rangle_G$  iff once you remove  $Z$  from the graph  $X$  and  $Y$  are not connected
- Can we completely capture probabilistic independencies by the notion of separation in a graph?
- Example: 2 coins and a bell.

# Graphoids vs Undirected graphs

## Graphoids: Conditional Independence

Symmetry:  $I(X,Z,Y) \rightarrow I(Y,Z,X)$

Decomposition:  $I(X,Z,YW) \rightarrow I(X,Z,Y)$  and  $I(X,Z,W)$

Weak union:  $I(X,Z,YW) \rightarrow I(X,ZW,Y)$

Contraction:  $I(X,Z,Y)$  and  $I(X,ZY,W) \rightarrow I(X,Z,YW)$

Intersection:  $I(X,ZY,W)$  and  $I(X,ZW,Y) \rightarrow I(X,Z,YW)$

## Separation in Graphs

Symmetry:  $I(X,Z,Y) \rightarrow I(Y,Z,X)$

Decomposition:  $I(X,Z,YW) \rightarrow I(X,Z,Y)$  and  $I(X,Z,W)$

Intersection:  $I(X,ZW,Y)$  and  $I(X,ZY,W) \rightarrow I(X,Z,YW)$

Strong union:  $I(X,Z,Y) \rightarrow I(X,ZW, Y)$

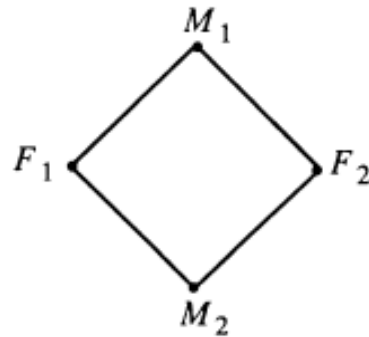
Transitivity:  $I(X,Z,Y) \rightarrow$  exists  $t$  s.t.  $I(X,Z,t)$  or  $I(t,Z,Y)$

See Pearl's book

# Markov Networks

An undirected graph  $G$  which is a minimal I-map of a probability distribution  $P$ , namely deleting any edge destroys its i-mappness relative to (undirected) separation, is called a **Markov network of  $P$** .

## MARKOV NETWORK AS A KNOWLEDGE BASE



**How can we construct a probability Distribution that will have all these independencies?**

Figure 3.2. An undirected graph representing interactions among four individuals.

### QUANTIFYING THE LINKS

- If couple  $(M_1, F_2)$  meet less frequently than the couple  $(M_1, F_1)$ , then the first link should be weaker than the second
- The model must be consistent, complete and a Markov field of  $G$ .
- Arbitrary specification of  $P(M_1, F_1)$ ,  $P(F_1, M_2)$ ,  $P(M_2, F_2)$ , and  $P(F_2, M_1)$  might lead to inconsistencies.
- If we specify the pairwise probabilities of only three pairs, incompleteness will result.

# Markov Random Field (MRF)

- A safe method (called *Gibbs' potential*) for constructing a complete and consistent quantitative model while preserving the dependency structure of an arbitrary graph  $G$ .
  1. Identify the cliques<sup>†</sup> of  $G$ , namely, the largest subgraphs whose nodes are all adjacent to each other.
  2. For each clique  $C_i$ , assign a nonnegative compatibility function  $g_i(c_i)$ , which measures the relative degree of compatibility associated with the value assignment  $c_i$  to the variables included in  $C_i$ .
  3. Form the product  $\prod_i g_i(c_i)$  of the compatibility functions over all the cliques.
  4. Normalize the product over all possible value combinations of the variables in the system

$$P(x_1, \dots, x_n) = K \prod_i g_i(c_i), \quad (3.13)$$

**So, How do we learn Markov networks From data?** where

$$K = \left[ \sum_{x_1, \dots, x_n} \prod_i g_i(c_i) \right]^{-1}.$$

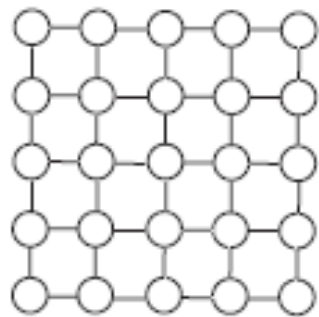
slides3b 276 2024

<sup>†</sup> We use the term *clique* for the more common term *maximal clique*.

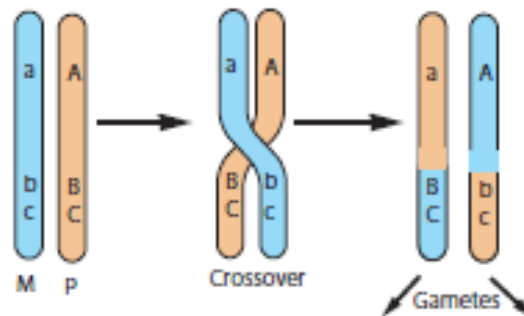


# Sample Applications for Graphical Models

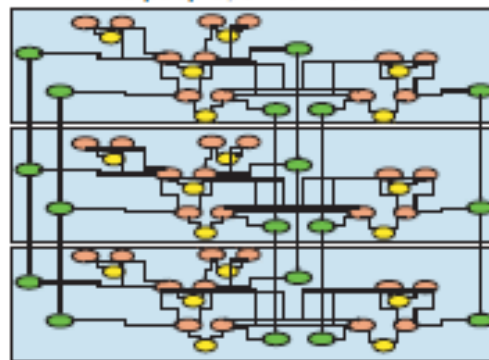
## Computer Vision



## Genetic Linkage



6 people, 3 markers



## Sensor Networks



Figure 1: Application areas and graphical models used to represent their respective systems: (a) Finding correspondences between images, including depth estimation from stereo; (b) Genetic linkage analysis and pedigree data; (c) Understanding patterns of behavior in sensor measurements using spatio-temporal models.

# Outline

- Bayesian networks and queries
- Building Bayesian Networks
  - Medical diagnosis
  - Circuit diagnosis
  - Probabilistic decoding
  - Commonsense reasoning
  - Linkage analysis

# Outline

- Bayesian networks and queries
- Building Bayesian Networks
  - Medical diagnosis
  - Circuit diagnosis
  - Probabilistic decoding
  - Commonsense reasoning
  - Linkage analysis

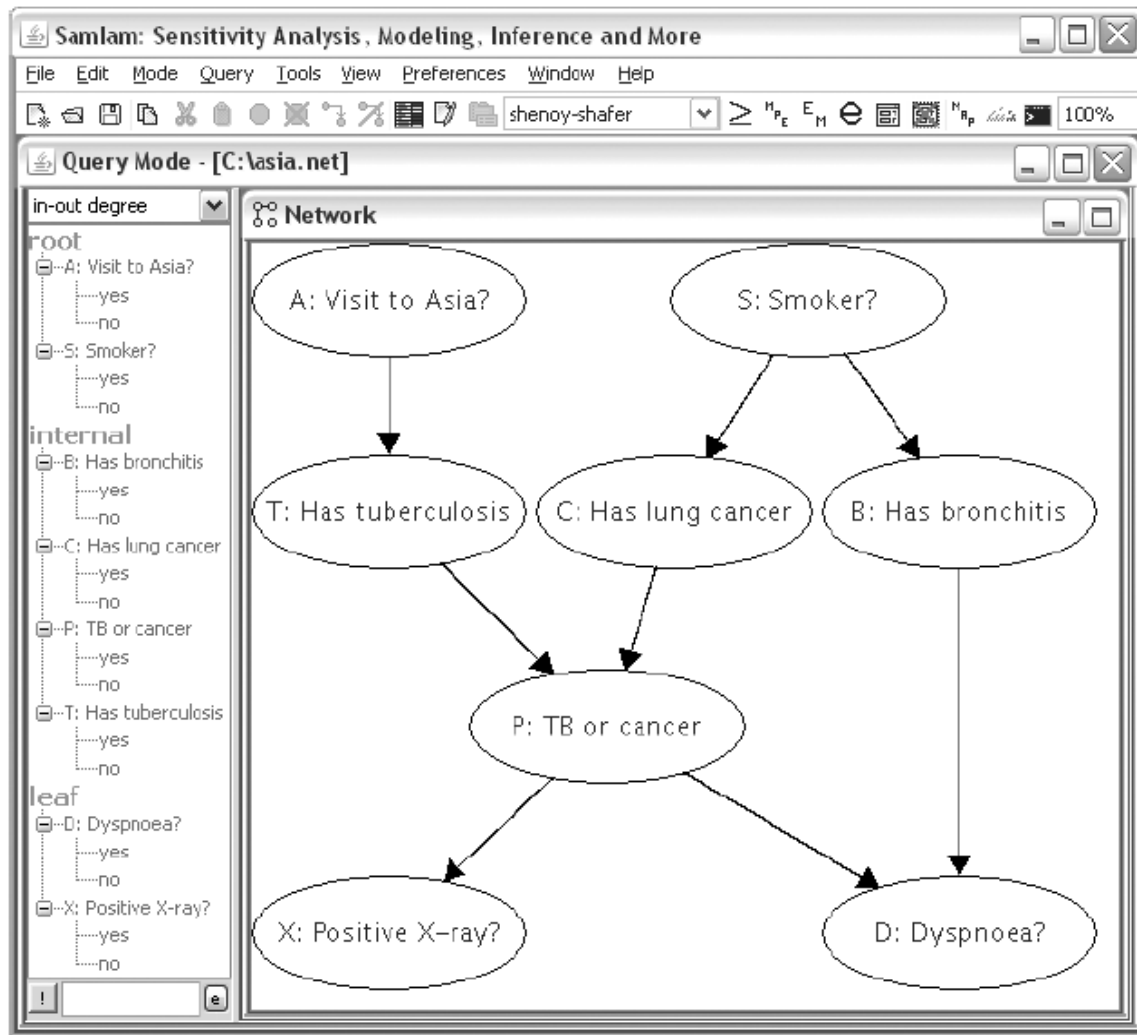
# Outline

The construction of a Bayesian network involves three major steps:

- Identify relevant variables and their possible values.
- Build the network structure by connecting variables into DAG.
- Define the CPT for each network variable.

**Queries:** Different queries may be relevant for different scenarios

# Reasoning with Bayesian Networks



The network **Asia** will be used as a running example. Screenshot from Samlam.

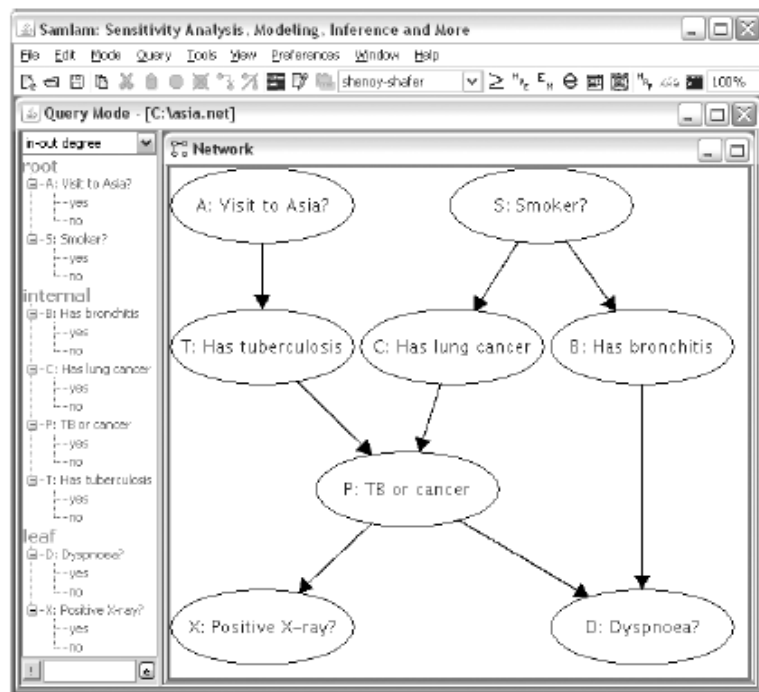
<http://reasoning.cs.ucla.edu/samiam>

Samlam available at <http://reasoning.cs.ucla.edu/samiam/>.

For other tools (e.g., GeNie/Smile) see class page

# Query: Probability of Evidence

Probability of some variable instantiation  $\mathbf{e}$ ,  $\Pr(\mathbf{e})$ .



Probability that the patient has a positive X-ray, but no dyspnoea,  $\Pr(X = \text{yes}, D = \text{no})$ , about 3.96%. Computed by Samlam.

The variables  $\mathbf{E} = \{X, D\}$  are called **evidence variables**. The query  $\Pr(\mathbf{e})$  is known as a **probability-of-evidence**.

**Other type of evidence:** We may want to know the probability that the patient has either a positive X-ray or dyspnoea,  $X = \text{yes}$  **or**  $D = \text{yes}$ .

# Query: Prior and Posterior Marginals

## Prior Marginals

Given a joint probability distribution  $\Pr(x_1, \dots, x_n)$ , the **marginal distribution**  $\Pr(x_1, \dots, x_m)$ ,  $m \leq n$ , is defined as follows:

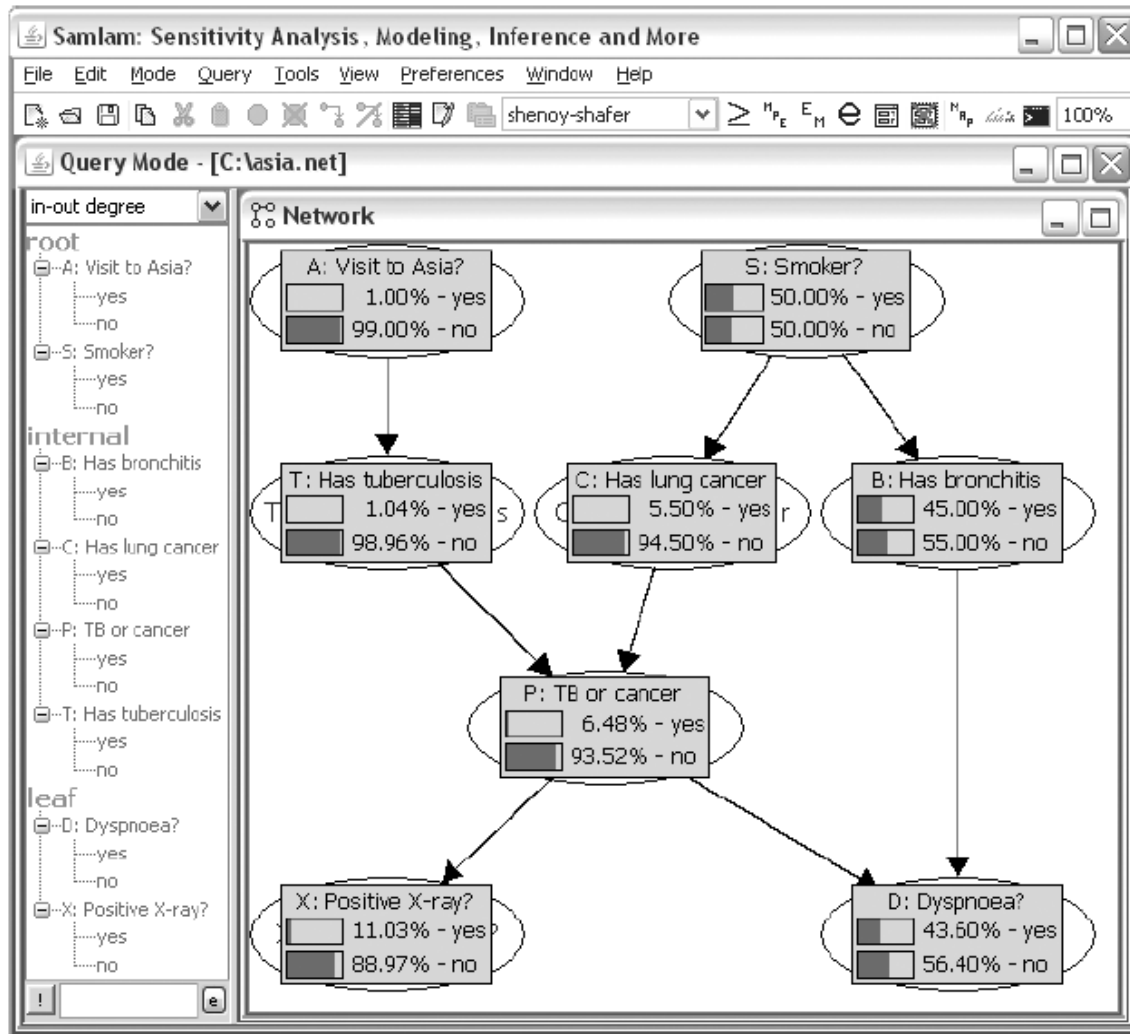
$$\Pr(x_1, \dots, x_m) = \sum_{x_{m+1}, \dots, x_n} \Pr(x_1, \dots, x_n).$$

The marginal distribution can be viewed as a **projection** of the joint distribution on the smaller set of variables  $X_1, \dots, X_m$ .

## Posterior marginal given evidence $\mathbf{e}$

$$\Pr(x_1, \dots, x_m | \mathbf{e}) = \sum_{x_{m+1}, \dots, x_n} \Pr(x_1, \dots, x_n | \mathbf{e}).$$

# Prior Marginals in the Asia Network

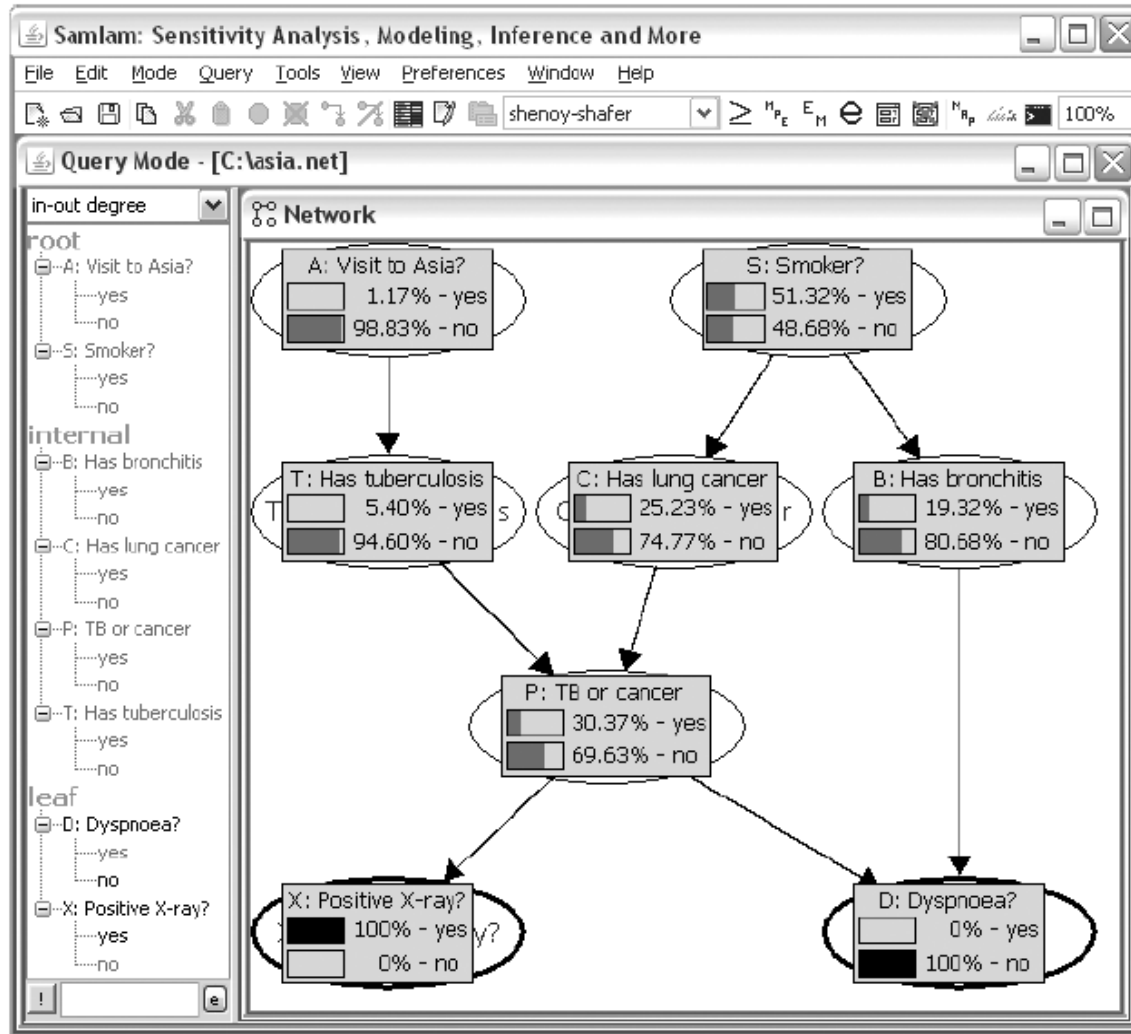


C= lung cancer

Prior marginal	
C	Pr(C)
yes	5.50%
no	94.50%



# Query: Posterior Marginals in the Asia Network



## Posterior marginal

C	Pr(C e)
yes	25.23%
no	74.77%

$e : X = \text{yes}, D = \text{no}$

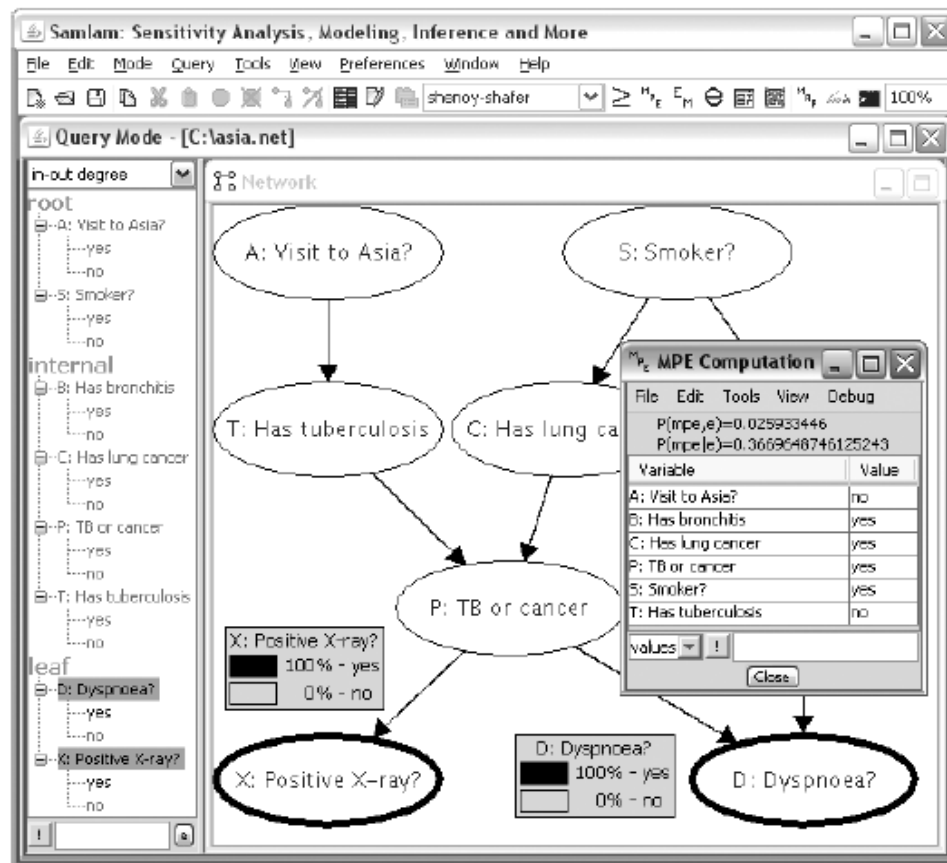
## Query: Most Probable Explanation (MPE)

Let  $X_1, \dots, X_n$  be all network variables, and  $\mathbf{e}$  be evidence. Identify an instantiation  $x_1, \dots, x_n$  that maximizes the probability  $\Pr(x_1, \dots, x_n | \mathbf{e})$ . Instantiation  $x_1, \dots, x_n$  is called a **most probable explanation** given evidence  $\mathbf{e}$ .

MPE cannot be obtained directly from posterior marginals.

If  $x_1, \dots, x_n$  is an instantiation obtained by choosing each value  $x_i$  so as to maximize the probability  $\Pr(x_i | \mathbf{e})$ , then  $x_1, \dots, x_n$  is not necessarily an MPE.

# Query: Most Probable Explanation (MPE)

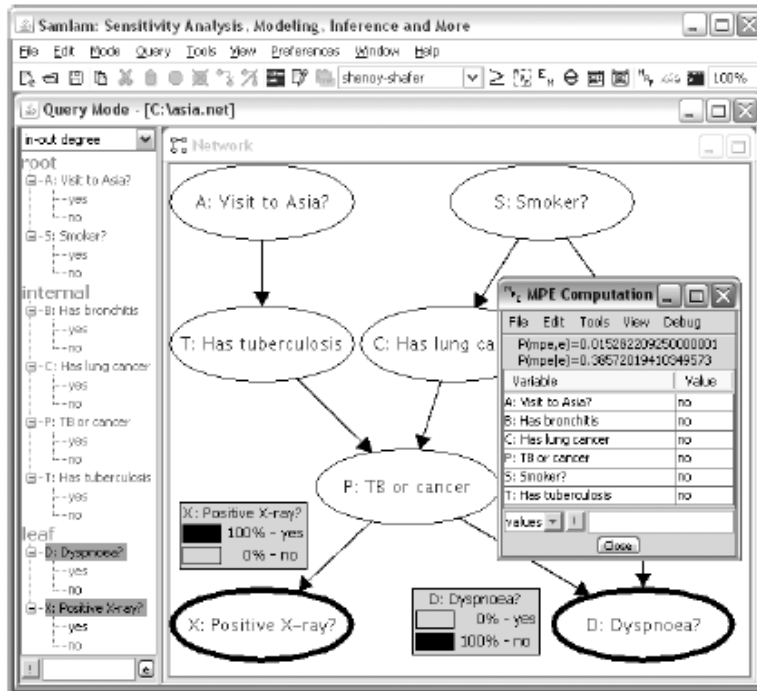


MPE given a positive X-ray and dyspnoea

A patient that made no visit to Asia; is a smoker; has lung cancer and bronchitis; but no tuberculosis.

MPE is also called MAP

# Query: Most Probable Explanation (MPE)



MPE given a positive X-ray and no dyspnoea ( $\approx 38.57\%$ )

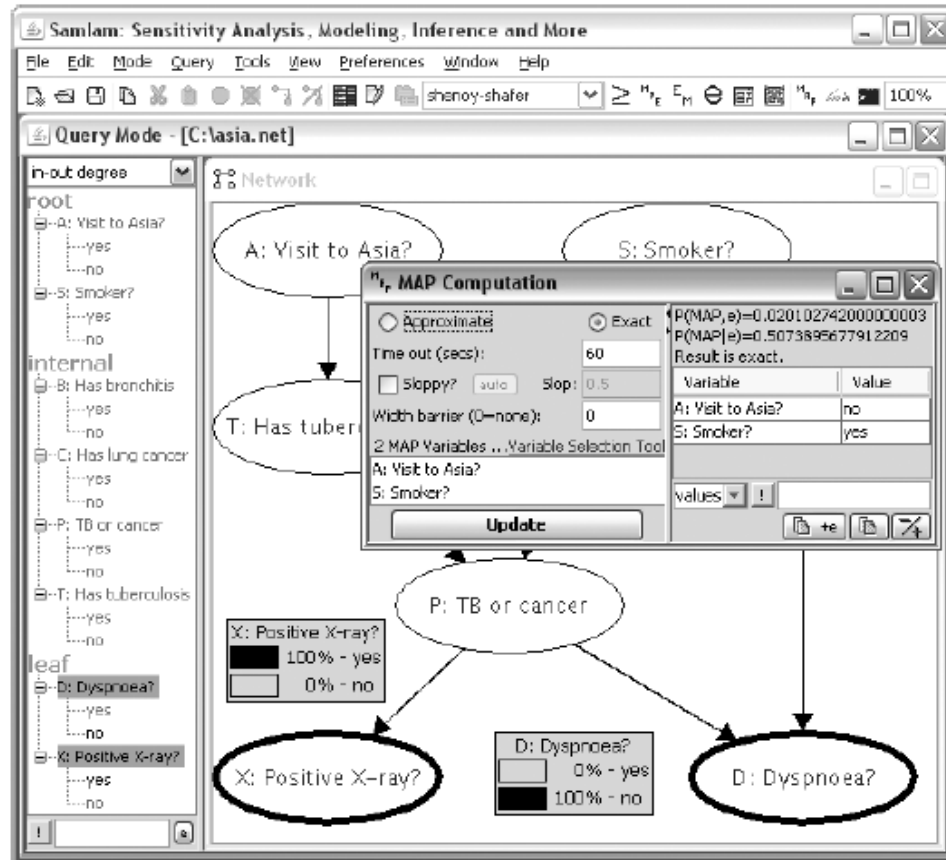
A patient that made no visit to Asia; is not a smoker; has no lung cancer, no bronchitis and no tuberculosis.

Choosing values with maximal probability, we get:

$\alpha$ :  $A = \text{no}$ ,  $S = \text{yes}$ ,  $T = \text{no}$ ,  $C = \text{no}$ ,  $B = \text{no}$ ,  $P = \text{no}$ ,  $X = \text{yes}$ ,  $D = \text{no}$ .

Probability  $\approx 20.03\%$  given evidence  $\mathbf{e}$ :  $X = \text{yes}$ ,  $D = \text{no}$ .

# Query: Maximum a Posteriori Hypothesis (MAP)



MAP variables

$M = \{A, S\}$  and  
evidence

$e : X = \text{yes}, D = \text{no}$

MAP is ~~A=no~~, ~~S=yes~~.

MAP has probability of  $\approx 50.74\%$  given the evidence.

**MAP is also called Marginal Map (MMAP)**

# Query: Maximum a Posteriori Hypothesis (MAP)

A common method for approximating MAP is to compute an MPE and then return the values it assigns to MAP variables. We say in this case that we are **projecting** the MPE on MAP variables.

Example

MPE  $\pi$

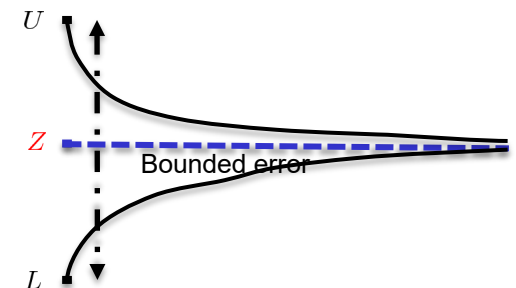
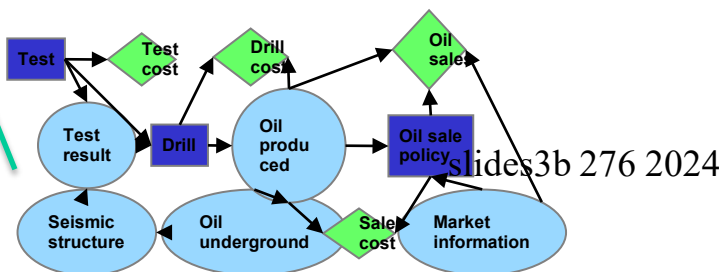
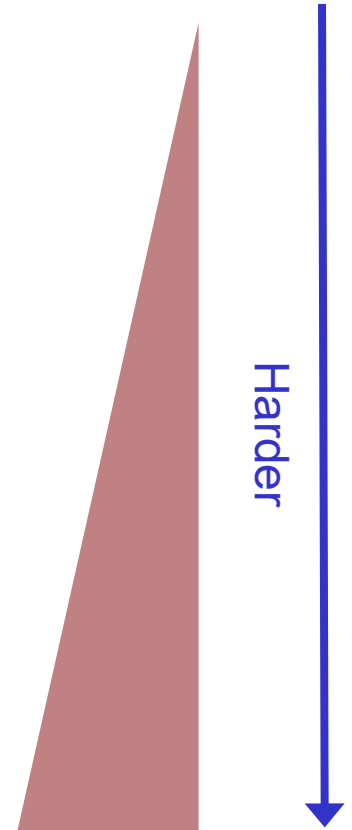
Is it correct?

MAP

# Probabilistic Reasoning Problems

Exact Algorithm: Bucket Elimination, Complexity  $e^{\text{tree-width}}$

<ul style="list-style-type: none"> <li>▶ <b>Max-Inference</b> (most likely config.)</li> </ul>	$f(\mathbf{x}^*) = \max_{\mathbf{x}} \prod_{\alpha} f_{\alpha}(\mathbf{x}_{\alpha})$
<ul style="list-style-type: none"> <li>▶ <b>Sum-Inference</b> (data likelihood)</li> </ul>	$Z = \sum_{\mathbf{x}} \prod_{\alpha} f_{\alpha}(\mathbf{x}_{\alpha})$
<ul style="list-style-type: none"> <li>▶ <b>Mixed-Inference</b> (optimal prediction)</li> </ul>	$f(\mathbf{x}_M^*) = \max_{\mathbf{x}_M} \sum_{\mathbf{x}_S} \prod_{\alpha} f_{\alpha}(\mathbf{x}_{\alpha})$
<ul style="list-style-type: none"> <li>▶ <b>Mixed-Inference</b> (maximum expected utility)</li> </ul>	$\text{MEU} = \max_{\Delta} \mathbb{E}_{P(\mathbf{X}, \mathbf{D})} [\sum_{U_i \in \mathbf{U}} U_i]$



# Modeling with Bayesian Networks

Bayesian networks will be constructed in three consecutive steps.

## Step 1

Define the network variables and their values.

- A **query variable** is one which we need to ask questions about, such as compute its posterior marginal.
- An **evidence variable** is one which we may need to assert evidence about.
- An **intermediary variable** is neither query nor evidence and is meant to aid the modeling process by detailing the relationship between evidence and query variables.

The distinction between query, evidence and intermediary variables is not a property of the Bayesian network, but of the task at hand.



# Modeling with Bayesian Networks

Bayesian networks will be constructed in three consecutive steps.

## Step 2

Define the network structure (edges).

Construction a minimal I-map of a  
distribution

We will be guided by a causal interpretation of network structure.

The determination of network structure will be reduced to answering the following question about each network variable  $X$ : what set of variables we regard as the direct causes of  $X$ ?

# Constructing a Bayesian Network for any Distribution $P$

**COROLLARY 3:** Given a probability distribution  $P(x_1, x_2, \dots, x_n)$  and any ordering  $d$  of the variables, the DAG created by designating as parents of  $X_i$  any minimal set  $\Pi_{X_i}$  of predecessors satisfying

$$P(x_i | \Pi_{X_i}) = P(x_i | x_1, \dots, x_{i-1}), \quad \Pi_{X_i} \subseteq \{X_1, X_2, \dots, X_{i-1}\} \quad (3.27)$$

is a Bayesian network of  $P$ .

- If  $P$  is strictly positive, then all of the parent sets are unique (see Theorem 4) and the Bayesian network is unique (given  $d$ ).

**COROLLARY 4:** Given a DAG  $D$  and a probability distribution  $P$ , a necessary and sufficient condition for  $D$  to be a Bayesian network of  $P$  is that each variable  $X$  be conditionally independent of all its non-descendants, given its parents  $\Pi_X$ , and that no proper subset of  $\Pi_X$  satisfy this condition.

Intuition: The causes of  $X$  can serve as the parents  
Ask: who does a variable listen to

# Modeling with Bayesian Networks

## Step 3

Define the network CPTs.

- CPTs can sometimes be determined completely from the problem statement by objective considerations.
- CPTs can be a reflection of subjective beliefs.
- CPTs can be estimated from data.

# Outline

- Bayesian networks and queries
- **Building Bayesian Networks**
  - Medical diagnosis
  - Circuit diagnosis
  - Probabilistic decoding
  - Commonsense reasoning
  - Linkage analysis
- Special representations of CPTs
  - Causal independence (noisy-or, noisy-and)
  - Decision trees

# Diagnosis I: Model from Expert

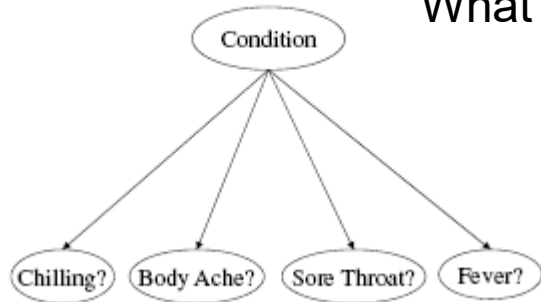
## Example

The flu is an acute disease characterized by fever, body aches and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils which leads to a sore throat and can be associated with fever.

Our goal here is to develop a Bayesian network to capture this knowledge and then use it to diagnose the condition of a patient suffering from some of the symptoms mentioned above.

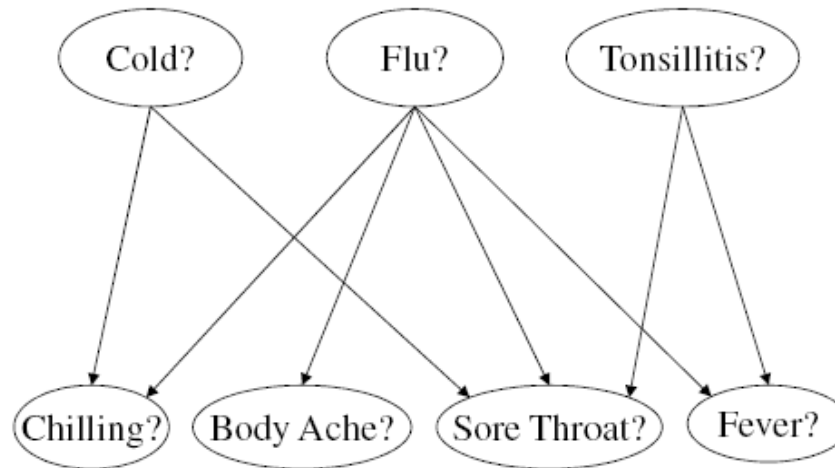
*Variables? Arcs? Try it.*

# Diagnosis I: Model from Expert



What about?

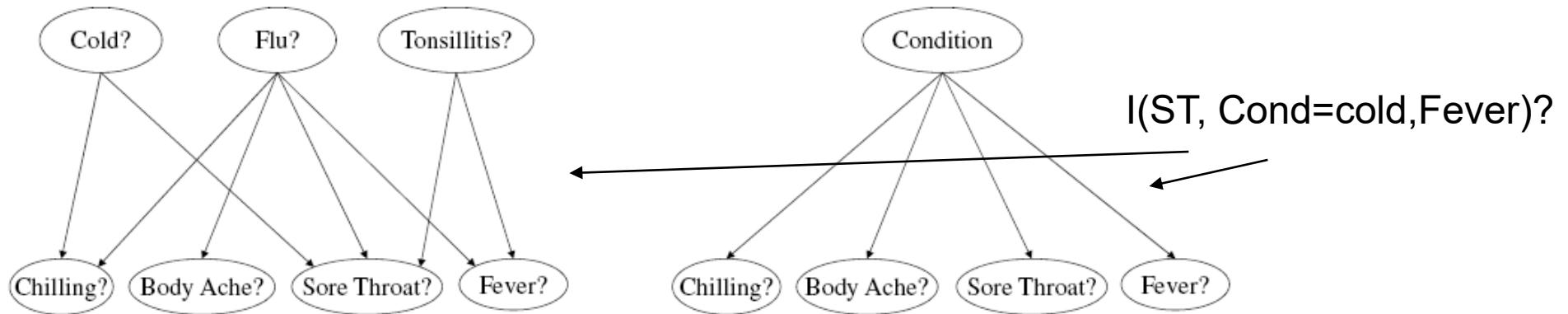
A naive Bayes structure has the following edges  $C \rightarrow A_1, \dots, C \rightarrow A_m$ , where  $C$  is called the class variable and  $A_1; \dots; A_m$  are called the attributes.



Variables are binary: values are either true or false. More refined information may suggest different degrees of body ache.

# Diagnosis I: Model from Expert

The naive Bayes structure commits to the **single-fault** assumption.



Suppose the patient is known to have a cold.

## Naive Bayes structure

Fever and sore throat become independent as they are d-separated by "Condition".

## Original structure

Fever may increase our belief in tonsillitis, which could then increase our belief in a sore throat.

# Diagnosis I: Learn the model from data

CPTs can be obtained from medical experts, who supply this information based on known medical statistics or subjective beliefs gained through practical experience.

CPTs can also be estimated from medical records of previous patients

<i>Case</i>	<i>Cold?</i>	<i>Flu?</i>	<i>Tonsillitis?</i>	<i>Chilling?</i>	<i>Bodyache?</i>	<i>Sorethroat?</i>	<i>Fever?</i>
1	true	false	?	true	false	false	false
2	false	true	false	true	true	false	true
3	?	?	true	false	?	true	false
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

? indicates the unavailability of corresponding data for that patient.



- Tools for Bayesian network inference can generate a network parameterization  $\Theta$ , which tries to maximize the probability of seeing the given cases.
- If each case is represented by event  $\mathbf{d}_i$ , such tools will generate a parametrization  $\Theta$  which leads to a probability distribution  $\text{Pr}$  that attempts to maximize:

$$\prod_{i=1}^N \text{Pr}(\mathbf{d}_i).$$

- Term  $\text{Pr}(\mathbf{d}_i)$  represents the probability of seeing the case  $i$ .
- The product represents the probability of seeing all  $N$  cases (assuming the cases are independent).

-

# Diagnosis II: Model from Expert

## Example

A few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is a scanning test which has a false positive of 1% and a false negative of 10%. The second is a blood test, which detects progesterone with a false positive of 10% and a false negative of 30%. The third test is a urine test, which also detects progesterone with a false positive of 10% and a false negative of 20%. The probability of a detectable progesterone level is 90% given pregnancy, and 1% given no pregnancy. The probability that insemination will impregnate a cow is 87%.

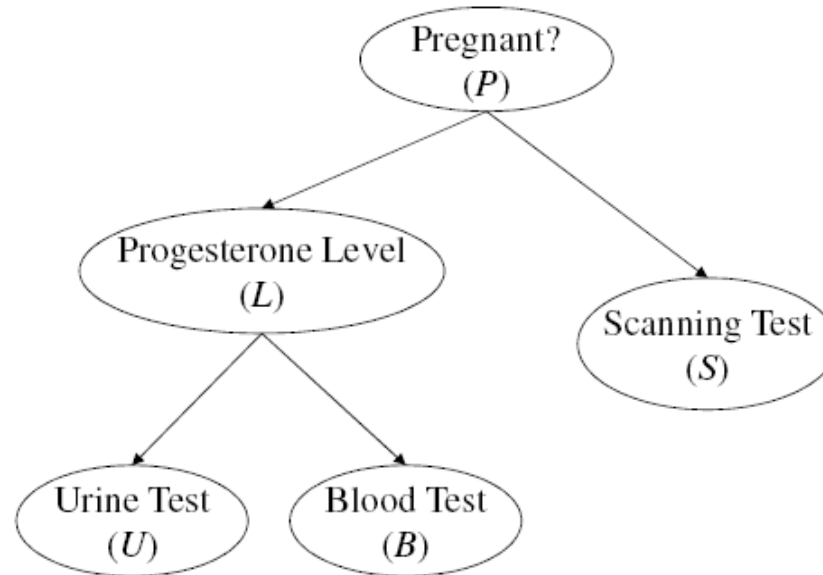
Our task here is to build a Bayesian network and use it to compute the probability of pregnancy given the results of some of these pregnancy tests.

Try it: Variables and values? Structure? CPTs?

slides 25-27 © 2024

# Diagnosis II: Model from Expert

Try with GeNie/Smile



$P$	$\theta_p$
yes	.87

$P$	$S$	$\theta_{s p}$
yes	-ve	.10
no	+ve	.01

$P$	$L$	$\theta_{l p}$
yes	undetectable	.10
no	detectable	.01

$L$	$B$	$\theta_{b l}$
detectable	-ve	.30
undetectable	+ve	.10

$L$	$U$	$\theta_{u l}$
detectable	-ve	.20
undetectable	+ve	.10

# Diagnosis II: Model from Expert

## Example

We inseminate a cow, wait for a few weeks, and then perform the three tests which all come out negative:

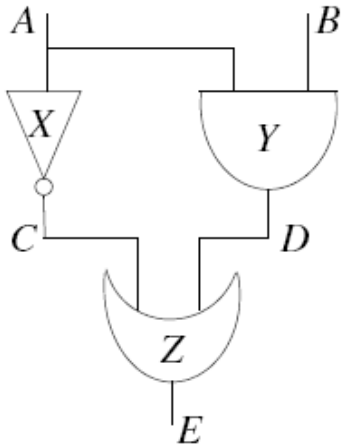
$$\mathbf{e}: S = -ve, B = -ve, U = -ve.$$

Posterior marginal for pregnancy given this evidence:

$P$	$\Pr(P \mathbf{e})$
yes	10.21%
no	89.79%

Probability of pregnancy is reduced from 87% to 10.21%, but still relatively high given that all three tests came out negative.

# Diagnosis III: Model from Design

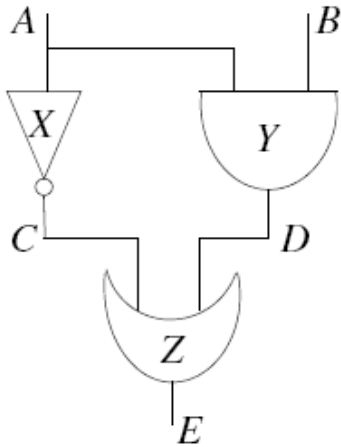


## Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

Try it: Variables? Values? Structure?

# Diagnosis III: Model from Design



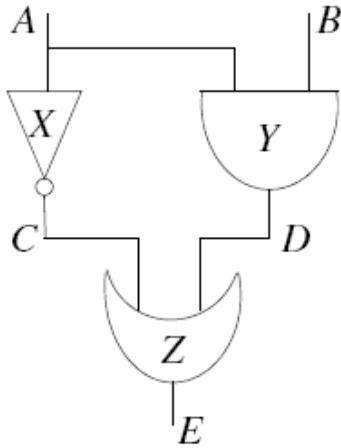
## Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

## Evidence variables

Primary inputs and output of the circuit,  $A$ ,  $B$  and  $E$ .

# Diagnosis III: Model from Design



## Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

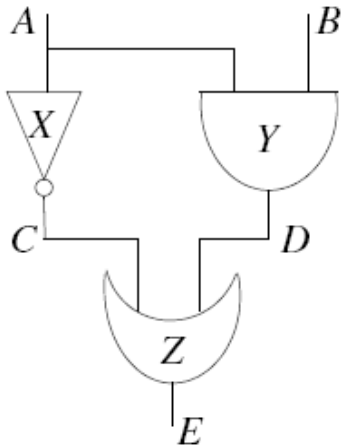
## Evidence variables

Primary inputs and output of the circuit,  $A$ ,  $B$  and  $E$ .

## Query variables

Health of components  $X$ ,  $Y$  and  $Z$ .

# Diagnosis III: Model from Design



## Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

## Evidence variables

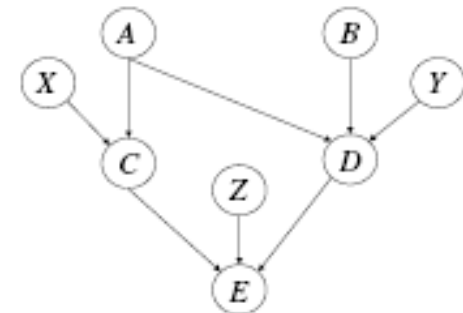
Primary inputs and output of the circuit,  $A$ ,  $B$  and  $E$ .

## Query variables

Health of components  $X$ ,  $Y$  and  $Z$ .

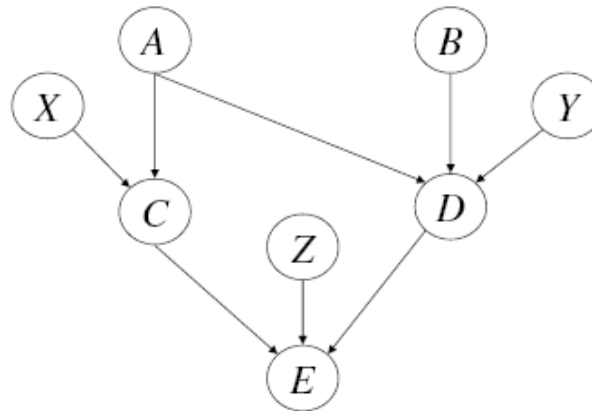
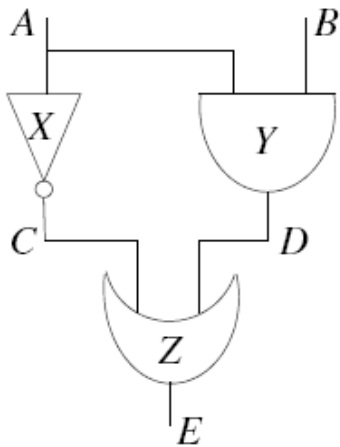
## Intermediary variables

Internal wires,  $C$  and  $D$ .





# Diagnosis III: Model from Design



Values of  
circuit wires:  
low or high

Health states: ok or faulty

faulty is too vague as a component may fail in a number of modes.

- **stuck-at-zero fault:** low output regardless of gate inputs.
- **stuck-at-one fault:** high output regardless of gate inputs.
- **input-output-short fault:** inverter shorts input to its output.

Fault modes demand more when specifying the CPTs.

# Diagnosis III: Model from Design

## Three classes of CPTs

- primary inputs ( $A, B$ )
- gate outputs ( $C, D, E$ )
- component health ( $X, Y, Z$ )

## CPTs for health variables depend on their values

$X$	$\theta_x$	$X$	$\theta_x$
ok	.99	ok	.99
faulty	.01	stuckat0	.005
		stuckat1	.005

Need to know the probabilities of various fault modes.

# Diagnosis III: Model from Design

CPTs for component outputs determined from functionality.

## Example

	$A$	$X$	$C$	$\theta_{c a,x}$
CPT for inverter $X$ .	high	ok	high	0
	low	ok	high	1
	high	stuckat0	high	0
	low	stuckat0	high	0
	high	stuckat1	high	1
	low	stuckat1	high	1

# Diagnosis III: Model from Design

CPTs for component outputs determined from functionality.

## Example

CPT for inverter X.

A	X	C	$\theta_{c a,x}$
high	ok	high	0
low	ok	high	1
high	stuckat0	high	0
low	stuckat0	high	0
high	stuckat1	high	1
low	stuckat1	high	1

If we do not represent health states:

A	X	C	$\theta_{c a,x}$
high	ok	high	0
low	ok	high	1
high	faulty	high	?
low	faulty	high	?

Common to use a probability of .50 in this case.

# A Diagnosis Example

## Example

Given test vector  $\mathbf{e}$ :  $A = \text{high}$ ,  $B = \text{high}$ ,  $E = \text{low}$ , compute MAP over health variables  $X$ ,  $Y$  and  $Z$ .

# A Diagnosis Example

## Example

Given test vector  $\mathbf{e}$ :  $A=\text{high}$ ,  $B=\text{high}$ ,  $E=\text{low}$ , compute MAP over health variables  $X$ ,  $Y$  and  $Z$ .

Network with fault modes gives two MAP instantiations:

MAP given $\mathbf{e}$	$X$	$Y$	$Z$	
	ok	stuckat0	ok	each probability $\approx 49.4\%$
	ok	ok	stuckat0	

# A Diagnosis Example

## Example

Given test vector  $\mathbf{e}$ :  $A = \text{high}$ ,  $B = \text{high}$ ,  $E = \text{low}$ , compute MAP over health variables  $X$ ,  $Y$  and  $Z$ .

Network with fault modes gives two MAP instantiations:

MAP given $\mathbf{e}$	$X$	$Y$	$Z$	
	ok	stuckat0	ok	each probability $\approx 49.4\%$
	ok	ok	stuckat0	

Network with no fault modes gives two MAP instantiations:

MAP given $\mathbf{e}$	$X$	$Y$	$Z$	
	ok	faulty	ok	each probability $\approx 49.4\%$
	ok	ok	faulty	

# Integrating Time

Suppose we have two test vectors instead of only one.



# Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

$A'$ ,  $B'$  and  $E'$

# Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

$A'$ ,  $B'$  and  $E'$

Additional intermediary variables

$C'$  and  $D'$

# Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

$A'$ ,  $B'$  and  $E'$

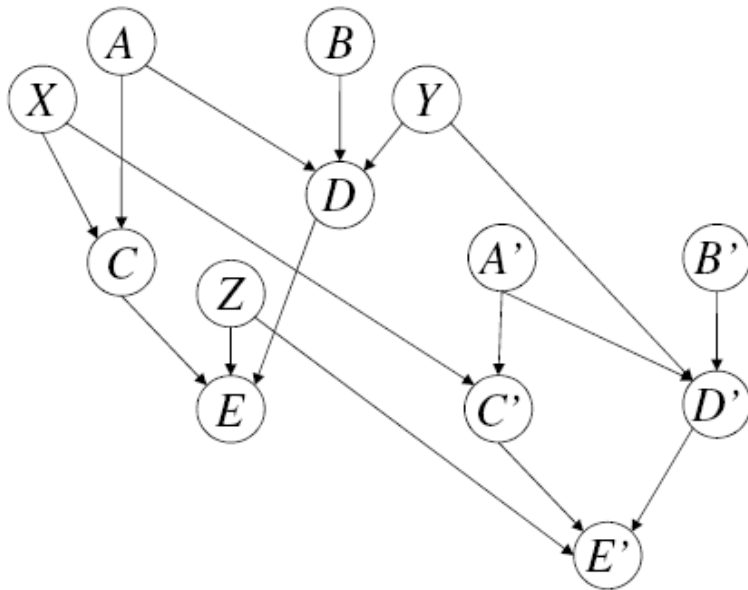
Additional intermediary variables

$C'$  and  $D'$

Additional health variables on whether we allow intermittent faults

If health of a component can change from one test to another, we need additional health variables  $X'$ ,  $Y'$ , and  $Z'$ . Otherwise, the original health variables are sufficient.

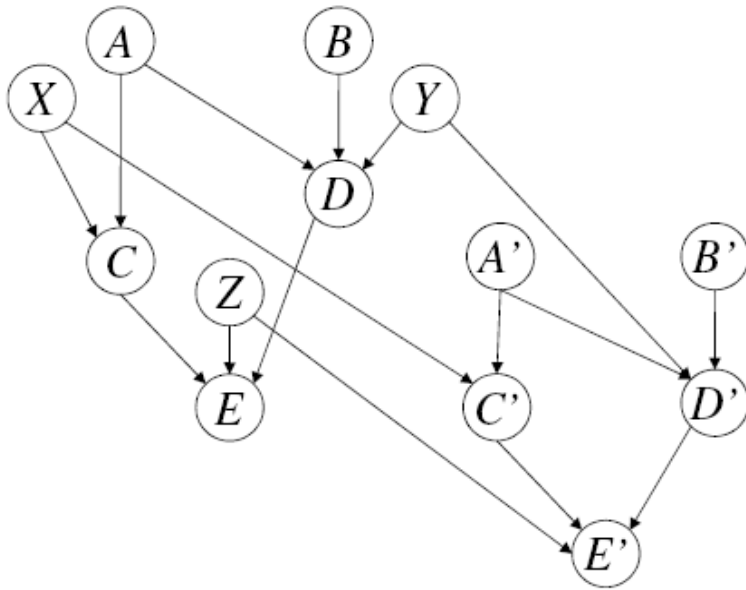
# Integrating Time: No Intermittent Faults



## Two test vectors

$e$  :  $A = \text{high}$ ,  $B = \text{high}$ ,  $E = \text{low}$   
 $e'$  :  $A = \text{low}$ ,  $B = \text{low}$ ,  $E = \text{low}$ .

# Integrating Time: No Intermittent Faults



## Two test vectors

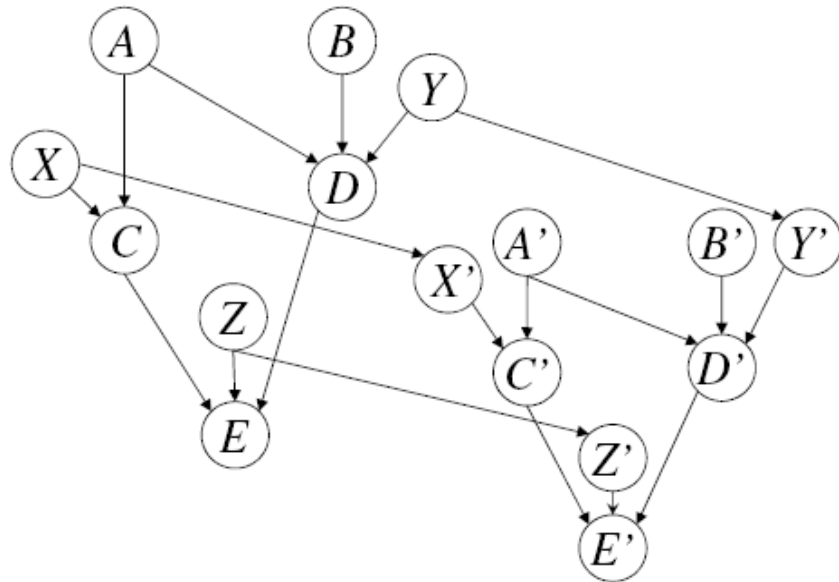
$e$  :  $A = \text{high}, B = \text{high}, E = \text{low}$   
 $e'$  :  $A = \text{low}, B = \text{low}, E = \text{low}$ .

## MAP using second structure

MAP given $e, e'$	$X$	$Y$	$Z$
	ok	ok	faulty

with probability  $\approx 97.53\%$

# Integrating Time: Intermittent Faults



Dynamic Bayesian network (DBN)

Two test vectors

$e$ :  $A = \text{high}$ ,  $B = \text{high}$ ,  $E = \text{low}$   
 $e'$ :  $A = \text{low}$ ,  $B = \text{low}$ ,  $E = \text{low}$ .

Persistence model for the health of component  $X$

$X$	$X'$	$\theta_{x' x}$	
ok	ok	.99	
ok	faulty	.01	healthy component becomes faulty
faulty	ok	.001	faulty component becomes healthy
faulty	faulty	.999	

Read on your own

# Commonsense reasoning

When SamBot goes home at night, he wants to know if his family is home before he tries the doors.

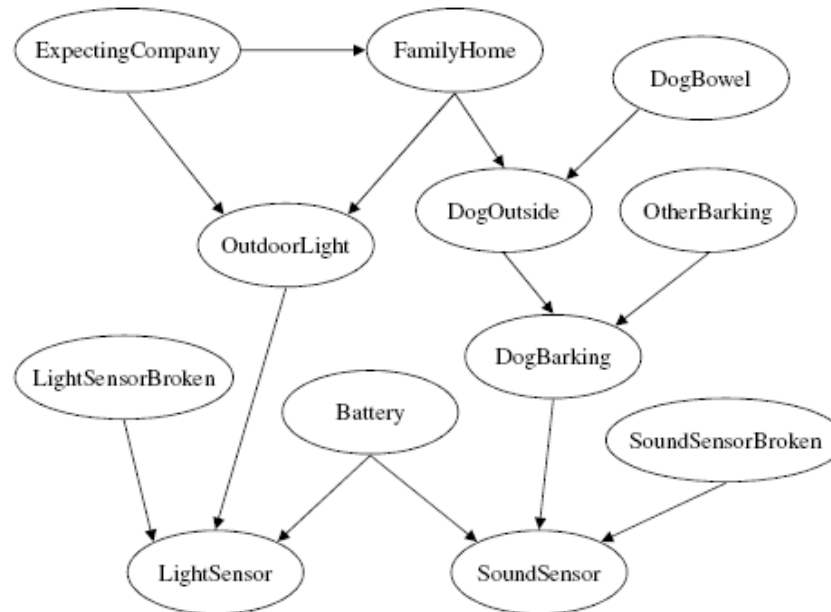
Often when SamBot's wife leaves the house she turns on an outdoor light. However, she sometimes turns on this light if she is expecting a guest.

Also, SamBot's family has a dog. When nobody is home, the dog is in the back yard. The same is true if the dog has bowel trouble.

If the dog is in the back yard, SamBot will probably hear her barking, but sometimes he can be confused by other dogs barking.

SamBot is equipped with two sensors: a light-sensor for detecting outdoor lights and a sound-sensor for detecting the barking of dogs. Both of these sensors are not completely reliable and can break. Moreover, they both require SamBot's battery to be in good condition.

# Commonsense Knowledge



## Parameters based on a combination of sources

- **Statistical information** such as reliabilities of sensors and battery.
- **Subjective beliefs** relating to how often the wife goes out, guests are expected, the dog has bowel trouble, etc.
- **Objective beliefs** regarding the functionality of sensors.



# Genetic Linkage Analysis

## A pedigree

is useful in reasoning about heritable characteristics which are determined by **genes**, where different genes are responsible for the expression of different characteristics.

# Genetic Linkage Analysis

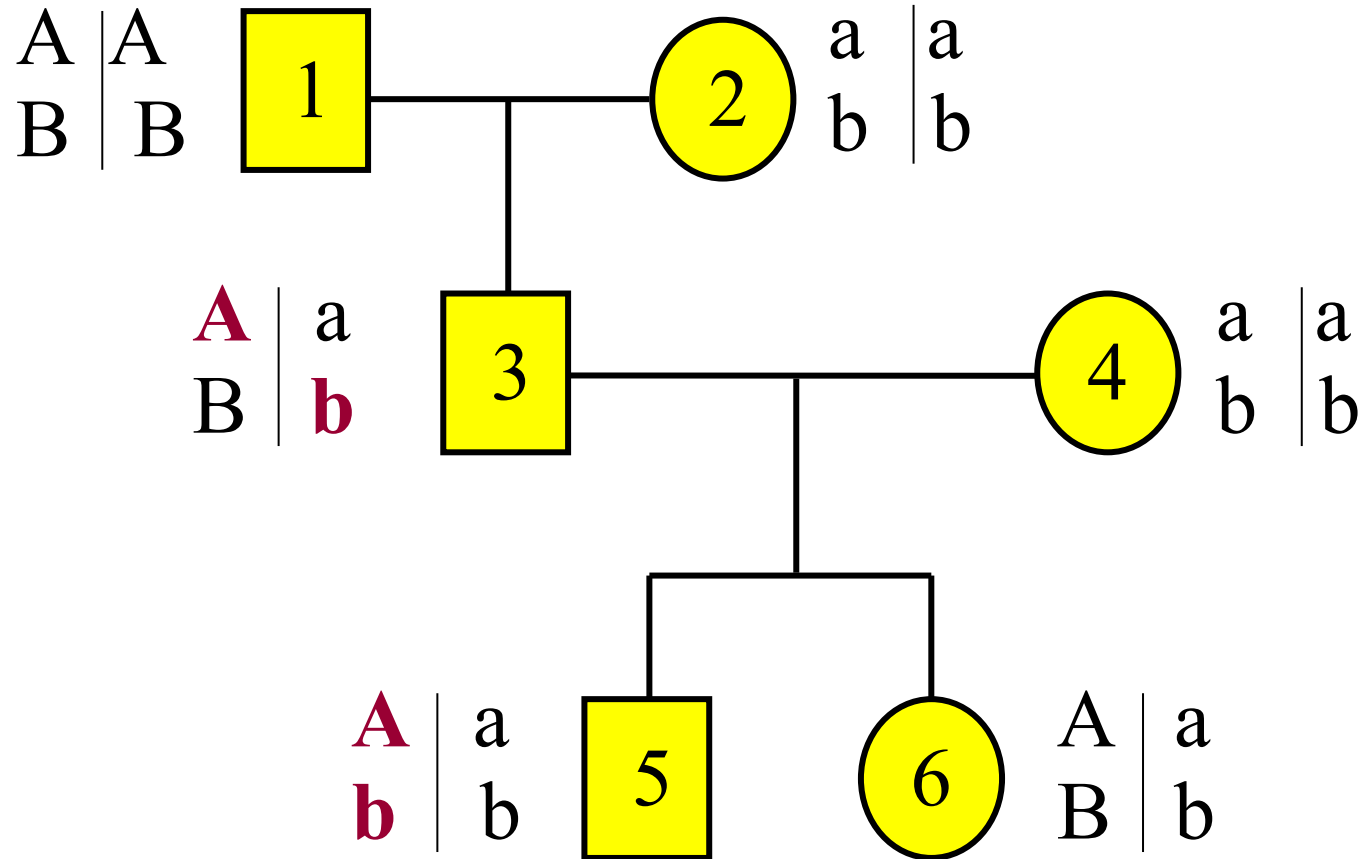
## A pedigree

is useful in reasoning about heritable characteristics which are determined by **genes**, where different genes are responsible for the expression of different characteristics.

## A gene

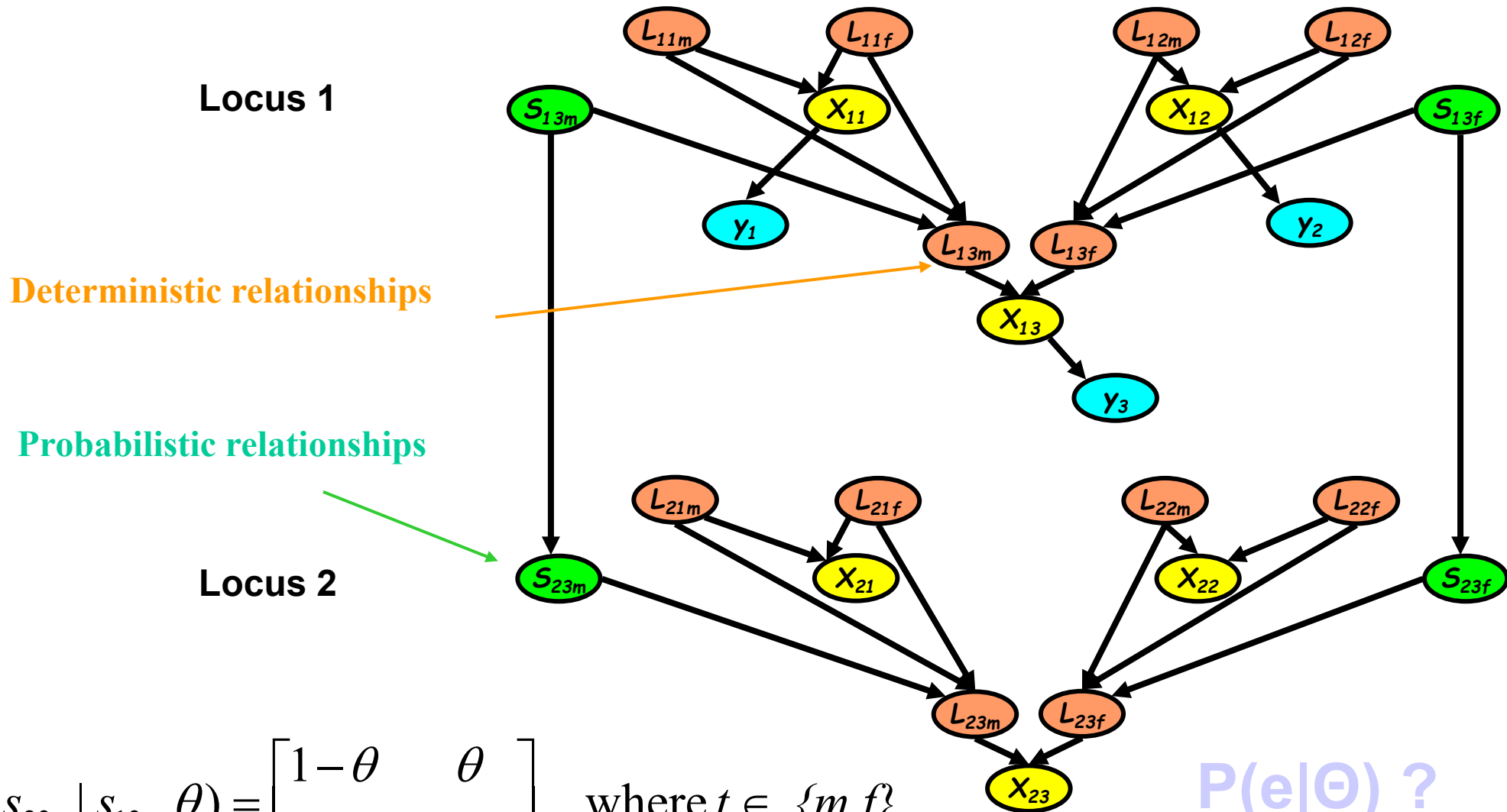
may occur in different states called **alleles**. Each individual carries two alleles of each gene, one received from their mother and the other from their father. The alleles of an individual are called the **genotype**, while the heritable characteristic expressed by these alleles (such as hair color, blood type, etc) are called the **phenotype** of the individual.

# Two Loci Inheritance



Recombinant

# Bayesian Network for Recombination



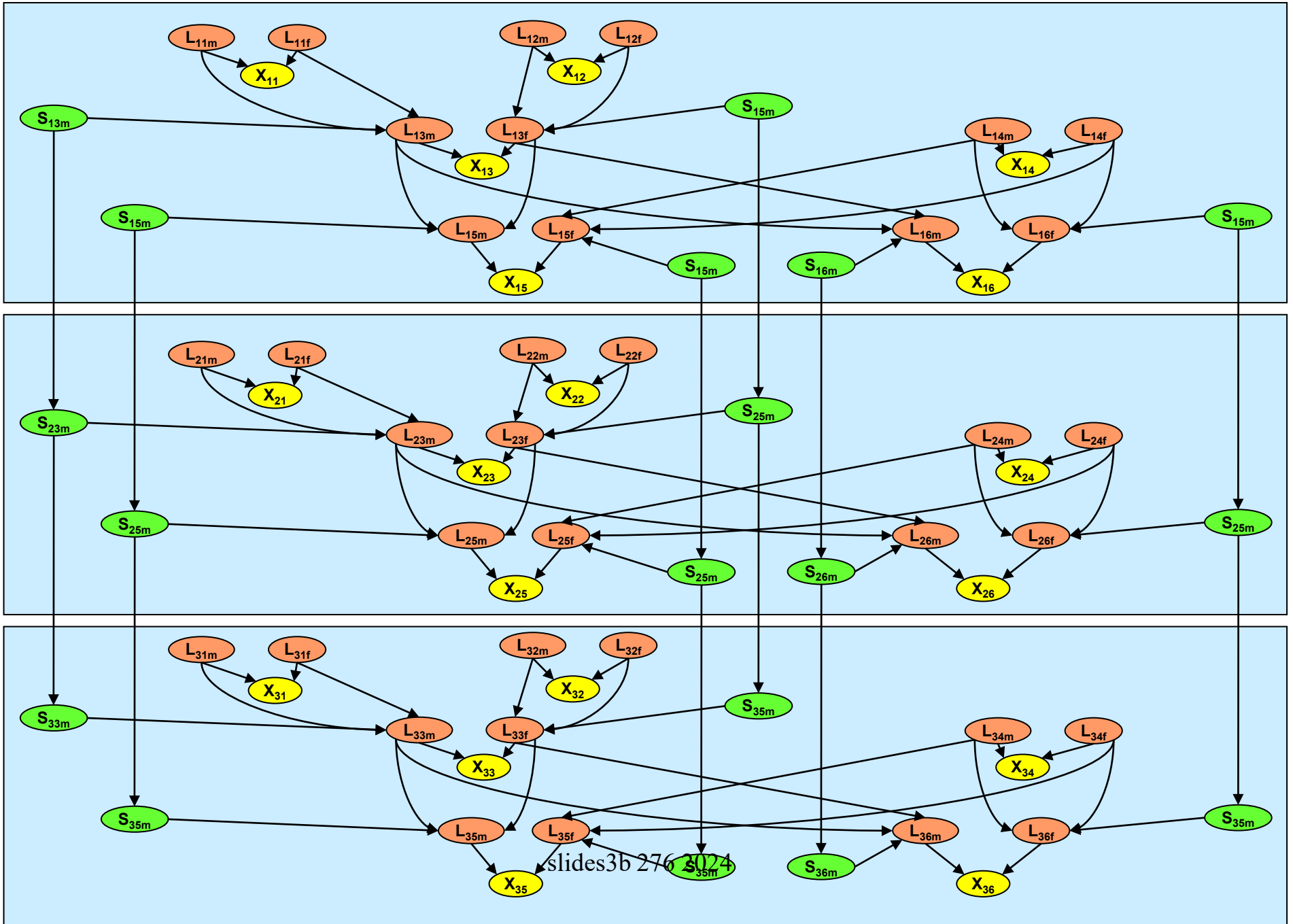
$$P(s_{23t} | s_{13t}, \theta) = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix}$$

where  $t \in \{m, f\}$

slides3b 276 2024

$P(e|\theta)$  ?

# Linkage analysis: 6 people, 3 markers



# Outline

- Bayesian networks and queries
- Building Bayesian Networks
  - Medical diagnosis
  - Circuit diagnosis
  - Probabilistic decoding
  - Commonsense reasoning
  - Linkage analysis