M.I. Jaime Alfonso Reyes ´Cortés

BAYESIAN NETWORKS

RED BAYESIANA

- Las redes Bayesianas son estructuras gráficas para representar las relaciones probabilísticas entre un gran número de variables. Permiten hacer inferencias probabilísticas con dichas variables
- Es un grafo dirigido acíclico (DAG) conexo más una distribución de probabilidad sobre sus variables

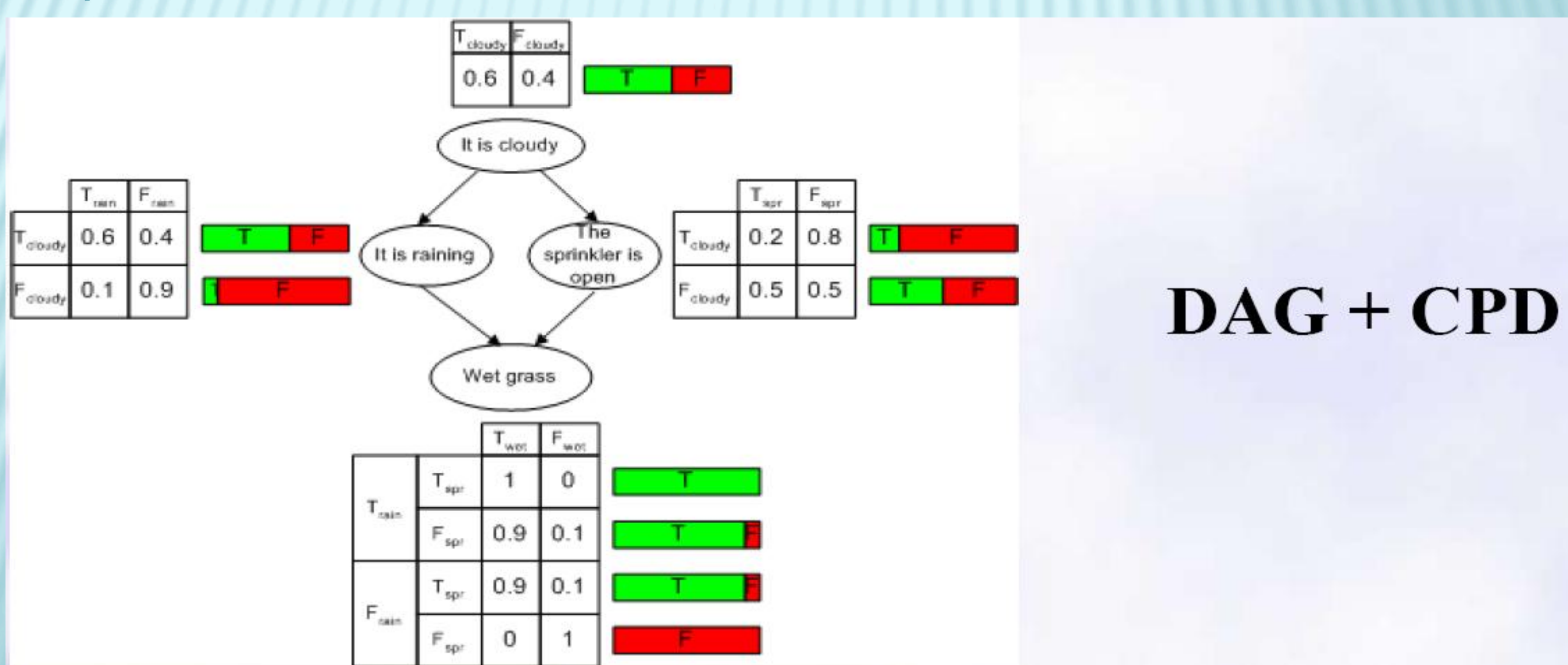


Figura 1. Red Bayesiana

PROPIEDADES

- En el DAG cada vértice puede contener un conjunto de alternativas mutuamente excluyentes y exhaustivas.
- Por tanto, la principal ventaja de las redes causales es que todas las alternativas pueden ser almacenadas en un vértice.
- Cada vértice representa una variable proposicional en lugar de una sola proposición.
- Una variable proposicional es una variable cuyo valor puede ser uno de un conjunto de alternativas mutuamente excluyentes y exhaustivas.

PROPIEDADES

- La dirección de un arco puede ser arbitraria, sin embargo, frecuentemente proviene desde una causa percibida hacia un efecto percibido, de aquí su nombre de causal
- Es necesario almacenar una probabilidad a priori para cada valor proposicional de cada nodo raíz y una probabilidad condicional para cada nodo no raíz con base en los valores proposicionales provenientes de sus padres . A partir de estos valores es posible determinar la nueva probabilidad condicional de cada valor proposicional de cada nodo dado cualquier otro nodo que haya sido instanciado para uno de sus valores

VENTAJAS DE LA REDES BAYESIANAS

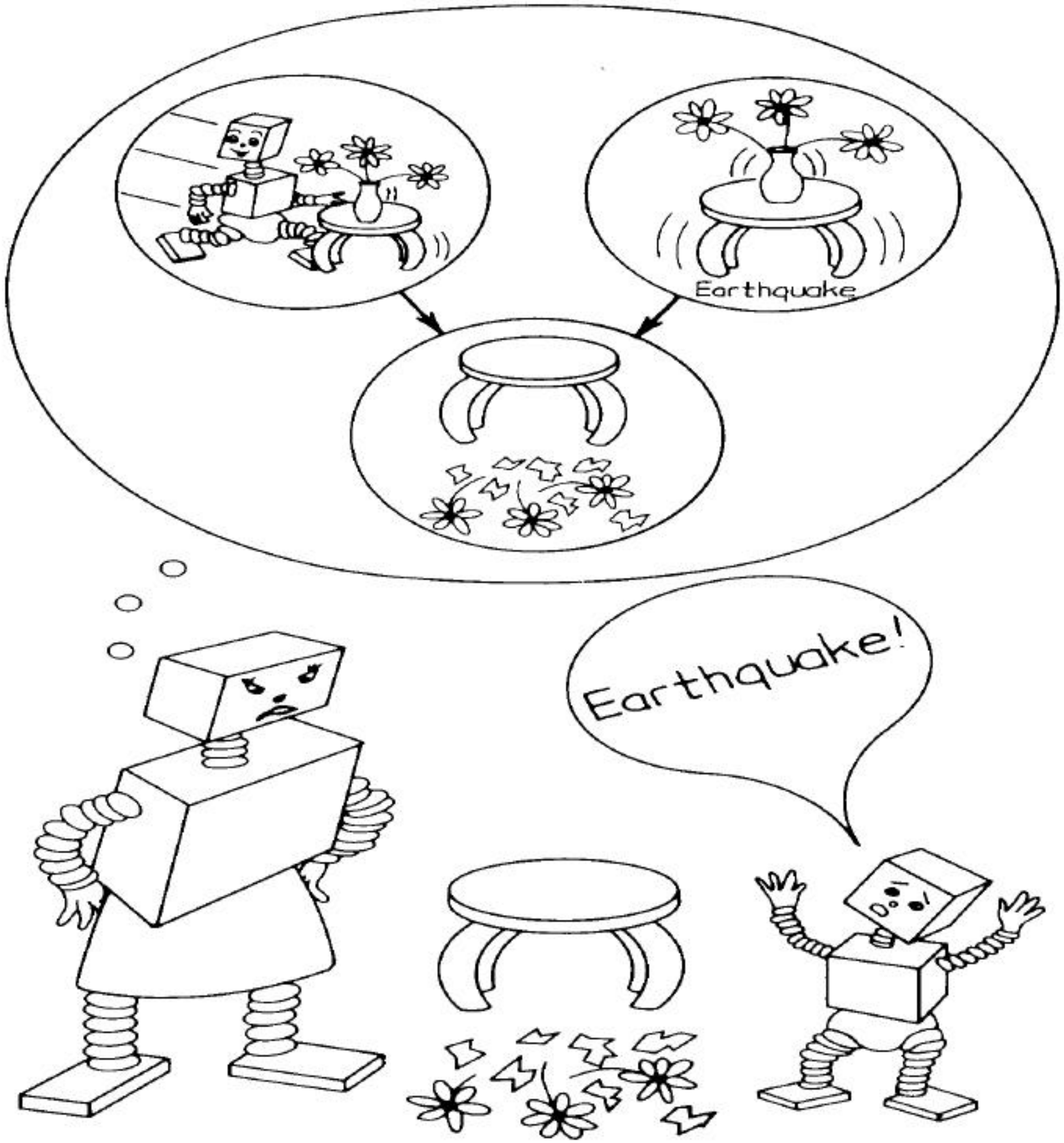
- La potencia explicativa de una interfaz de usuario inteligente se basa en : **transparencia**, o la habilidad de comprender cómo trabaja un sistema; y **flexibilidad**, o la habilidad de la interfaz para adaptarse a una gran variedad de interacciones con el usuario final.
- Las redes Bayesiana destacan en ambos aspectos.
- Primero, las redes Bayesianas son altamente intuitivas y gráficas. Nos permiten modelar factores o causas en un diagrama de red. Como son gráficas, se pueden observar fácilmente los efectos de una nueva evidencia en los nodos del diagrama causal.Por lo tanto, permiten un alto grado de transparencia en el proceso de razonamiento.
- Segundo, Las redes Bayesianas son altamente flexibles. La nueva evidencia puede ser introducida en cualquier orden (es decir, los nodos pueden ser instanciados en cualquier orden) y la red bayesiana se actualizará apropiadamente.
- Además las redes Bayesianas harán predicciones basadas en evidencia incompleta, lo que las hace robustas y no son propensas a caídas si alguna evidencia no se encuentra disponible.

VENTAJAS DE LA REDES BAYESIANAS

- Además , las redes bayesianas pueden razonar desde un efecto hacia una causa, no sólo desde la causa hacia el efecto. Por tanto, pueden llevar a cabo razonamiento hacia atrás, que no es posible en otras aproximaciones, incluyendo en la aproximación de factores de certidumbre.
- Otro beneficio de las redes bayesianas es que proporcionan una excelente manera de representar juicios subjetivos del experto.
- Las probabilidades a priori y condicionales especificadas inicialmente en las BN pueden representar la opinión subjetiva del experto o datos objetivos. Las BN procesarán cada tipo con el mismo desempeño, en comparación con otras aproximaaciones estadísticas, como el análisis de regresión que es una técnica altamente conducida por los datas y descuidad la opinión del experto.
- BN pueden incorporar la opinión del experto y la explicación causal. Son más aceptables para tareas que requieren de la predicción de eventos futuros.

TIPOS DE REDES

- Existen distintos tipos de redes Bayesianas:
  - Naive Bayes: Redes simples.
    - Forma de “V” => 2^n estados en el nodo inferior
  - DBNs: Redes Bayesianas Dinámicas
    - Cambian con el tiempo (t, t+1, t+2)
    - Lo pasado en t tiene relación con lo que suceda en t+1
  - Redes gaussianas: tienen distribución gaussiana. Para nodos con variables continuas
  - Cadenas de Markov: subconjunto de las redes Bayesianas



Probabilistic Reasoning in a Causal Network

Figura 2. Red Bayesiana simple

- Our goal in Bayesian modeling is, at least largely, to find the most accurate representation of a real system about which we may be receiving inconsistent expert advice, rather than finding ways of modeling the inconsistency itself.
- Bayesian networks provide a natural representation of probabilities which allow for (and take advantage of) any independencies that may hold, while not being limited to problems satisfying strong independence requirements.

- The combination of substantial increases in computer power with the Bayesian network’s ability to use any existing independencies to computational advantage make the approximations and restrictive assumptions of earlier uncertainty formalisms pointless. So we now turn to the main game: understanding and representing uncertainty with probabilities.

CONDITIONALIZATION

- *After applying Bayes' theorem to obtain P(h|e) adopt that as your posterior degree of belief in h*

$$Bel(h) = P(h|e)$$

- Conditionalization, in other words, advocates belief updating via probabilities conditional upon the available evidence. It identifies **posterior probability (the probability** function after incorporating the evidence, which we are writing Bel(.)) with **conditional probability**

EXPECTED UTILITY

- Generally, agents are able to assign utility (or, value) to the situations in which they find themselves. We know what we like, we know what we dislike, and we also know when we are experiencing neither of these. Given a general ability to order situations, and bets with definite probabilities of yielding particular situations,
- Frank Ramsey [231] demonstrated that we can identify particular utilities with each possible situation, yielding a **utility function**

EXPECTED UTILITY (CONT.)

- If we have a utility function $U(O_i|A)$ over every possible outcome of a particular action A we are contemplating, and if we have a probability for each such outcome $P(O_i|A)$ then we can compute the probability-weighted average utility for that action otherwise known as the **expected utility of the action:**

$$EU(A) = \sum_i U(O_i|A) \times P(O_i|A)$$

- It is commonly taken as axiomatic by Bayesians that agents ought to *maximize their expected utility. That is, when contemplating a number of alternative actions,* agents ought to decide to take that action which has the maximum expected utility.
- Utilities have behavioral consequences ess*entially: any agent who consistently ignores the putative* utility of an action or situation arguably does not have that utility.

FAIR BETS

- **Fair bets** are fair because their expected utility is zero.
- Suppose we are contemplating taking the fair bet B on proposition h for which we assign probability P(h) . Then the expected utility of the bet is:

$$EU(B) = U(h|B)P(h|B) + U(\neg h|B)P(\neg h|B)$$

- Sí $P(h|B) = P(h)$

FAIR BETS (CONT.)

$$EU(B) = U(h|B)P(h) + U(\neg h|B)(1 - P(h))$$

$$U(h|B) = 1 - P(h)$$

$$U(\neg h|B) = -P(h)$$

$$EU(B) = (1 - P(h))P(h) - P(h)(1 - P(h)) = 0$$

- Given that the bet has zero expected utility, the agent should be no more inclined to take the bet in favor of h than to take the opposite bet against h

BAYESIAN NETWORKS

- Bayesian networks (BNs) are graphical models for reasoning under uncertainty, where the nodes represent variables (discrete or continuous) and arcs represent direct connections between them.
- These direct connections are often causal connections.
- In addition, BNs model the quantitative strength of the connections between variables, allowing probabilistic beliefs about them to be updated automatically as new information becomes available.

BN BASICS

- A **Bayesian network is a graphical structure that allows us to represent and reason** about an uncertain domain.
- The nodes in a Bayesian network represent a set of random variables from the domain.
- A set of directed **arcs (or links) connects pairs of** nodes, representing the direct dependencies between variables.
- Assuming discrete variables, the strength of the relationship between variables is quantified by conditional probability distributions associated with each node.
- The only constraint on the arcs allowed in a BN is that there must not be any directed cycles
- There are a number of steps that a **knowledge engineer** must undertake when building a Bayesian network:

BN BASICS (CONT.)

- **Example problem: Lung cancer.** *A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer*

BN BASICS (CONT.) NODES AND VALUES

- First, the knowledge engineer must identify the variables of interest.
- what are the nodes to represent and what values can they take?
- The values should be both **mutually exclusive and exhaustive, which means that the variable** must take on exactly one of these values at a time. Common types of discrete nodes include:
  - Boolean nodes, which represent propositions, taking the binary values true (T) or false (F). E.g. the node *Cancer*

BN BASICS (CONT.) NODES AND VALUES

- Ordered values. For example, a node *Pollution might represent a patient's* pollution exposure and take the values {*low, medium, high}.*
- Integral values. For example, a node called *Age might represent a patient's age* and have possible values from 1 to 120.

- The trick is to choose values that represent the domain efficiently, but with enough detail to perform the reasoning required.

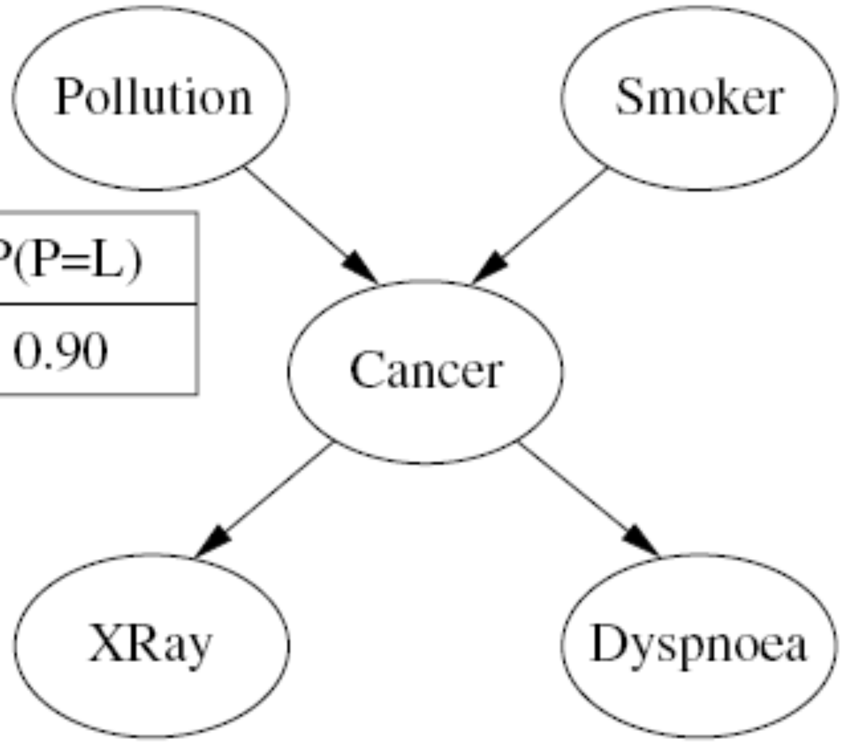
Preliminary choices of nodes and values for the lung cancer example

Node name	Type	Values
Pollution	Binary	{*low, high*}
Smoker	Boolean	{*T, F*}
Cancer	Boolean	{*T, F*}
Dyspnoea	Boolean	{*T, F*}
X-ray	Binary	{*pos, neg*}

BN BASICS (CONT.) STRUCTURE

- The structure, or topology, of the network should capture qualitative relationships between variables. In particular, two nodes should be connected directly if one affects or causes the other, with the arc indicating the direction of the effect.

BN BASICS (CONT.) STRUCTURE



P(S=T)
0.30

P(P=L)
0.90

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	P(X=pos|C)
T	0.90
F	0.20

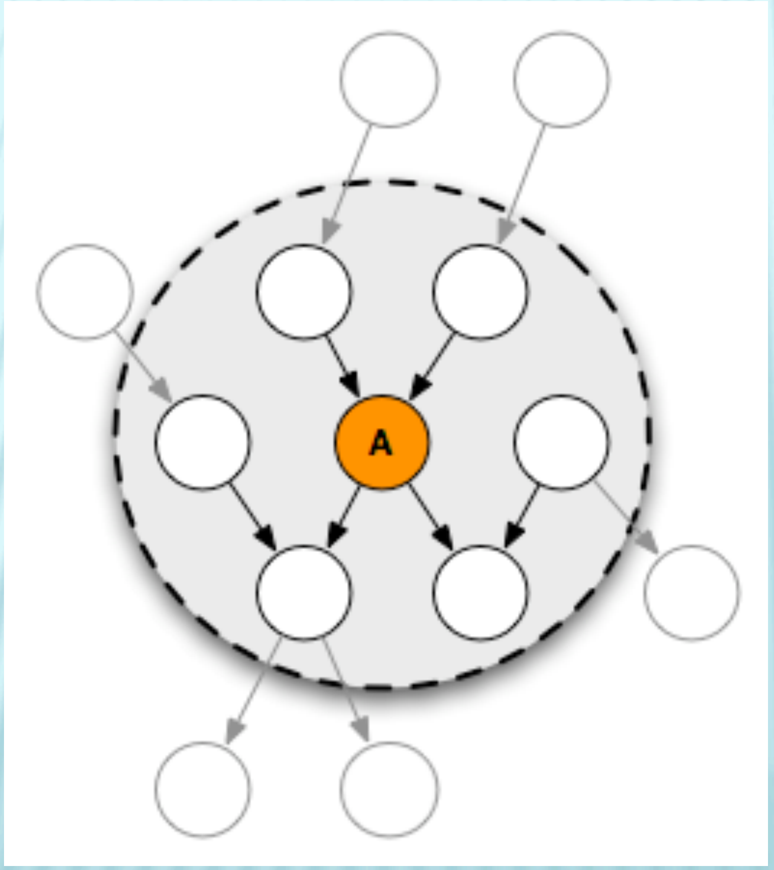
C	P(D=T|C)
T	0.65
F	0.30

BN BASICS (CONT.) STRUCTURE

- a node is a **parent of a child, if there is an arc from the former to the latter**
- if there is a directed chain of nodes, one node is an **ancestor of another if** it appears earlier in the chain, whereas a node is a **descendant of another node if it** comes later in the chain
- any node without parents is called a **root node,** while any node without children is called a **leaf node. Any other node (non-leaf** and non-root) is called an **intermediate node.**
- root nodes represent original causes, while leaf nodesrepresent final effects

BN BASICS (CONT.) STRUCTURE

- Markov blanket



BN BASICS (CONT.) CONDITIONAL PROBABILITIES

- To specify the probability distribution of a Bayesian network, one must provide two types of probabilities: (1) the prior probabilities of the root nodes (nodes with no parents) and (2) the conditional probabilities of all nonroot nodes given all combinations of its parent nodes.
- The next step is to quantify the relationships between connected nodes (specifying a conditional probability distribution for each node). Considering discrete variables this takes the form of a *conditional probability table (CPT)*

BN BASICS (CONT.) CONDITIONAL PROBABILITIES

- First, for each node we need to look at all the possible combinations of values of those parent nodes.
- Each such combination is called an **instantiation of the parent** set.
- For each distinct instantiation of parent node values, we need to specify the probability that the child will take each of its values.

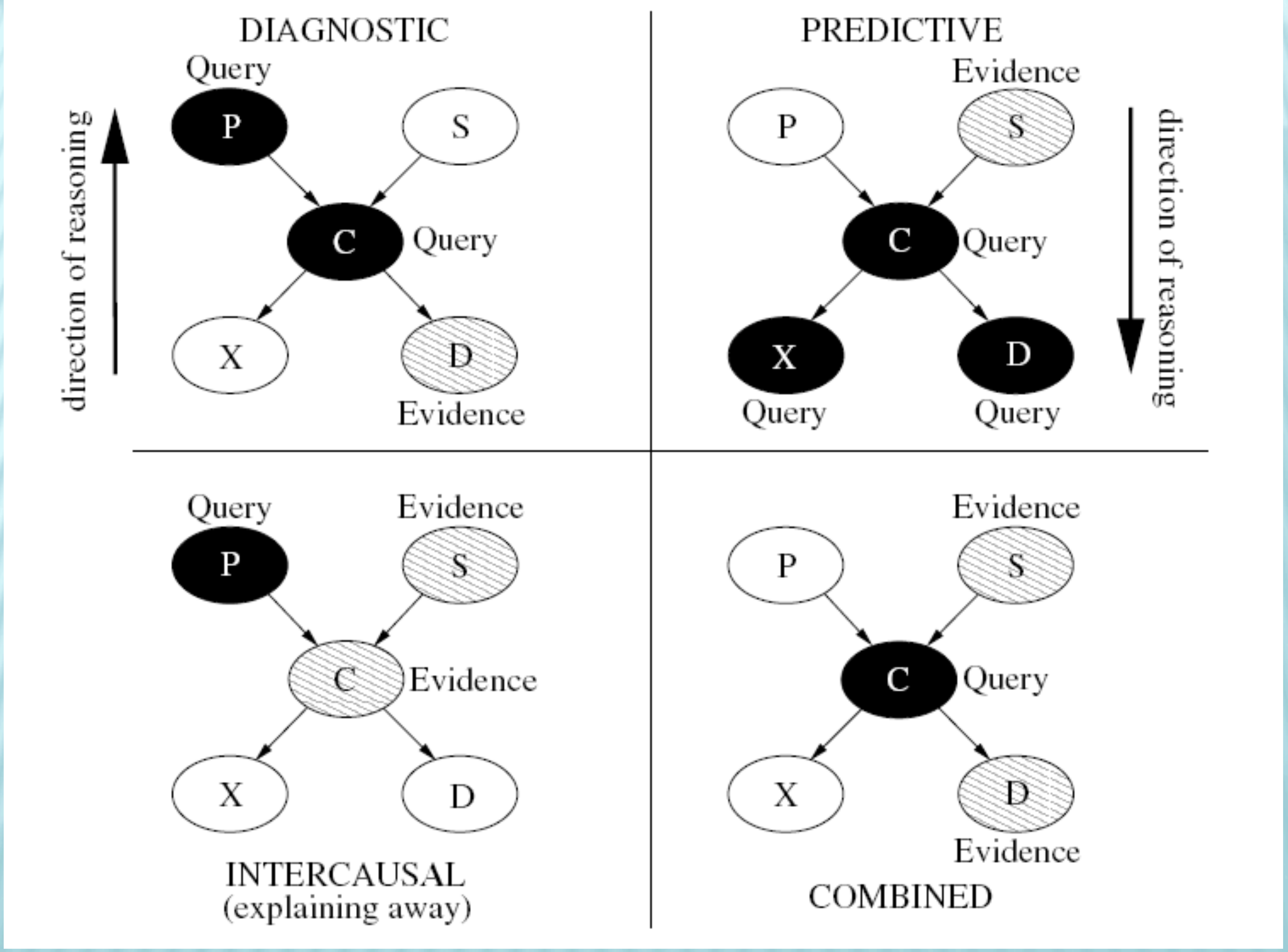
BN BASICS (CONT.) THE MARKOV PROPERTY

- there are no direct dependencies in the system being modeled which are not already explicitly shown via arcs
- Bayesian networks which have the Markov property are also called **Independence-maps (or, I-maps** for short), since every independence suggested by the lack of an arc is real in the system.
- every arc in a BN happens to correspond to a direct dependence in the system, then the BN is said to be a **Dependence-map (or, D-map for short). A BN** which is both an I-map and a D-map is said to be a **perfect map.**

REASONING WITH BAYESIAN NETWORKS

- The process of conditioning (also called **probability propagation or inference or belief updating)** is performed via a "flow of information" through the network
- In our probabilistic system, this becomes the task of computing the posterior probability distribution for a set of **query nodes, given values** for some **evidence (or observation) nodes**

REASONING WITH BN: TYPES OF REASONING



- **diagnostic reasoning, i.e.,** reasoning from symptoms to cause
- **predictive reasoning, reasoning from new information** about causes to new beliefs about effects, following the directions of the network arcs
- **intercausal reasoning** reasoning about the mutual causes of a common effect

Node P(S)=0.3	No Evidence	Reasoning Case Diagnostic D=T	Predictive S=T	Intercausal C=T	C=T S=T	Combined D=T S=T
Bel(P=high)	0.100	0.102	0.100	0.249	0.156	0.102
Bel(S=T)	0.300	0.307	1	0.825	1	1
Bel(C=T)	0.011	0.025	0.032	1	1	0.067
Bel(X=pos)	0.208	0.217	0.222	0.900	0.900	0.247
Bel(D=T)	0.304	1	0.311	0.650	0.650	1
P(S)=0.5						
Bel(P=high)	0.100	0.102	0.100	0.201	0.156	0.102
Bel(S=T)	0.500	0.508	1	0.917	1	1
Bel(C=T)	0.174	0.037	0.032	1	1	0.067
Bel(X=pos)	0.212	0.226	0.311	0.900	0.900	0.247
Bel(D=T)	0.306	1	0.222	0.650	0.650	1

REPRESENTING THE JOINT PROBABILITY DISTRIBUTION

- BNs are considered to be representations of joint probability distributions
- There is a fundamental assumption that there is a useful underlying structure to the problem being modeled that can be captured with a BN
- A BN gives a more compact representation than simply describing the probability of every joint instantiation of all variables

REPRESENTING THE JOINT PROBABILITY DISTRIBUTION

- Consider a BN containing the n nodes X1 to Xn. A particular value in the joint distribution is represented by

$$P(x_1, x_2, \ldots, x_n) = P(x_1) \times P(x_2|x_1) \ldots, \times P(x_n|x_1, \ldots, x_{n-1})$$
$$= \prod_i P(x_i|x_1, \ldots, x_{i-1})$$

- Recalling the structure of a BN implies that the value of a particular node is conditional *only on the values of its parent nodes, this reduces to*

$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|Parents(X_i))$$

$$Parents(X_i) \subseteq \{x_1, \ldots, x_{i-1}\}$$

$$
\begin{aligned}
P(X = pos \wedge D = T \wedge C = T \wedge P = low \wedge S = F)& \\
= P(X = pos | D = T, C = T, P = low, S = F)& \\
\times P(D = T | C = T, P = low, S = F)& \\
\times P(C = T | P = low, S = F) P(P = low | S = F) P(S = F)& \\
= P(X = pos | C = T) P(D = T | C = T) P(C = T | P = low, S = F)& \\
\times P(P = low) P(S = F)&
\end{aligned}
$$

PEARL'S NETWORK CONSTRUCTION ALGORITHM

- The condition that $Parents(X_i) \subseteq \{x_1, \ldots, x_{i-1}\}$
- allows us to construct a network from a given ordering of nodes using Pearl's network construction algorithm
- Furthermore, the resultant network will be a unique minimal I-map, assuming the probability distribution is positive.
- The construction algorithm simply processes each node in order, adding it to the existing network and adding arcs from a minimal set of parents such that the parent set renders the current node conditionally independent of every other node preceding it

Pearl's Network Construction Algorithm

1. *Choose the set of relevant variables* $\{X_i\}$ *that describe the domain.*
2. *Choose an ordering for the variables,* $< X_1, \ldots, X_n >$.
3. *While there are variables left:*
   (a) *Add the next variable* X_i *to the network.*
   (b) *Add arcs to the* X_i *node from some minimal set of nodes already in the net,* $Parents(X_i)$, *such that the following conditional independence property is satisfied:*

$$P(X_i|X'_1, \ldots, X'_m) = P(X_i|Parents(X_i))$$

   where $X'_1, \ldots, X'_m$ *are all the variables preceding* X_i *that are not in* $Parents(X_i)$.
   (c) *Define the CPT for* X_i.

COMPACTNESS AND NODE ORDERING

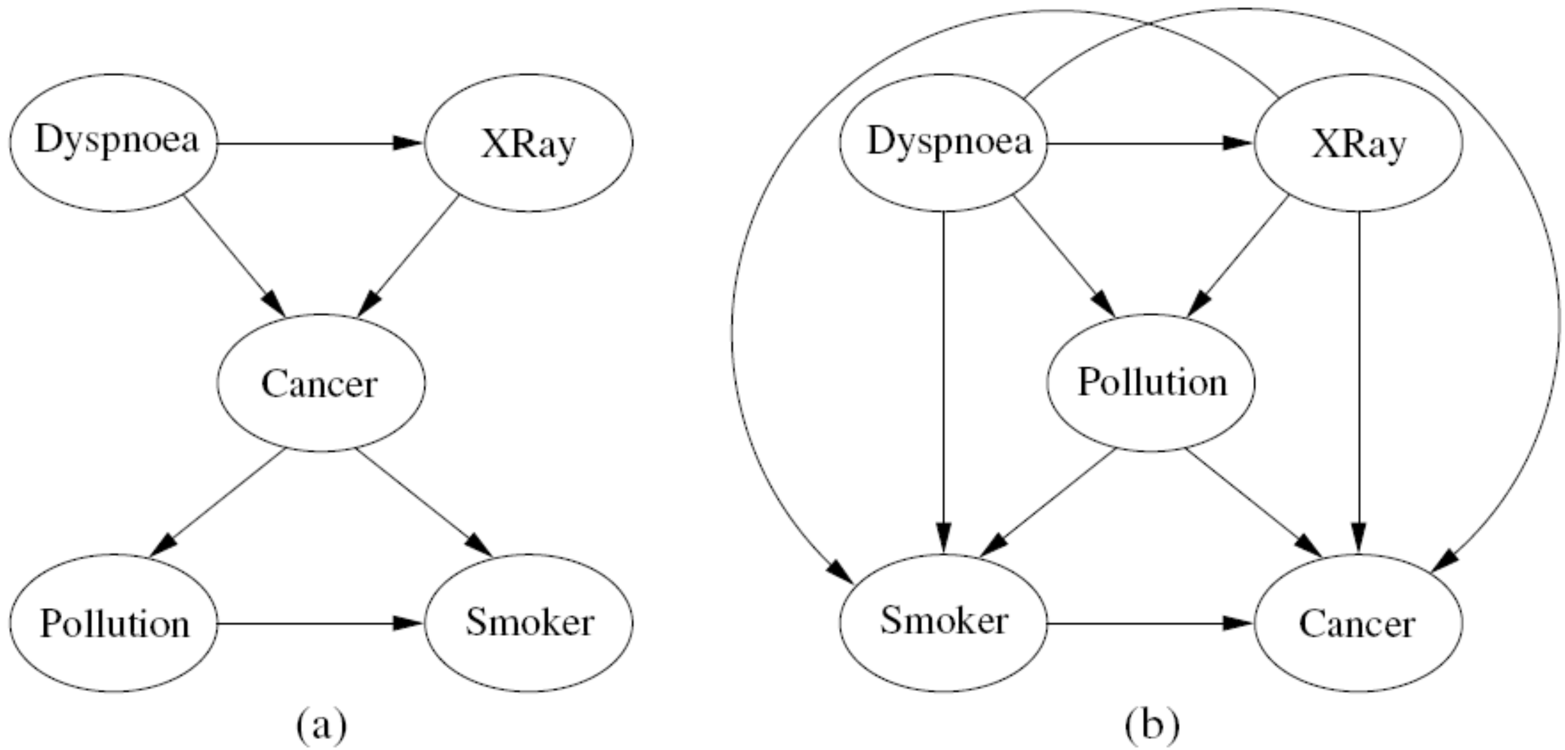


FIGURE 2.3

Alternative structures obtained using Pearl's network construction algorithm with orderings: (a) $< D, X, C, P, S >$; (b) $< D, X, P, S, C >$.

- It is desirable to build the most compact BN possible, for three reasons.
- First, the more compact the model, the more tractable it is. It will have fewer probability values requiring specification; it will occupy less computer memory; probability updates will be more computationally efficient.
- Second, overly dense networks fail to represent independencies explicitly.
- Third, overly dense networks fail to represent the *causal dependencies in the domain.*

- The optimal order is to add the root causes first, then the variable(s) they influence directly, and continue until leaves are reached. To understand *why, we need to consider the relation between probabilistic and causal* dependence.

CARACTERÍSTICAS DE LAS REDES BAYESIANAS

- The structure of a causal diagram can be used to propagate probabilities in a Bayesian network.
- One distinguishing feature of Bayesian networks is the ability to reason about problem structure and propagate probabilities accordingly.
- In this respect, Bayesian networks are really a form of model-based reasoning, a form of inferential reasoning in which problem solving is aided by knowledge of system structure because system behavior is determined from structure.

conditional independence as: $A \perp\!\!\!\perp C | B$

$$P(C|A \wedge B) = P(C|B)$$

$$P(A|C \wedge B) \neq P(A|C) \equiv \neg(A \perp\!\!\!\perp C | B)$$

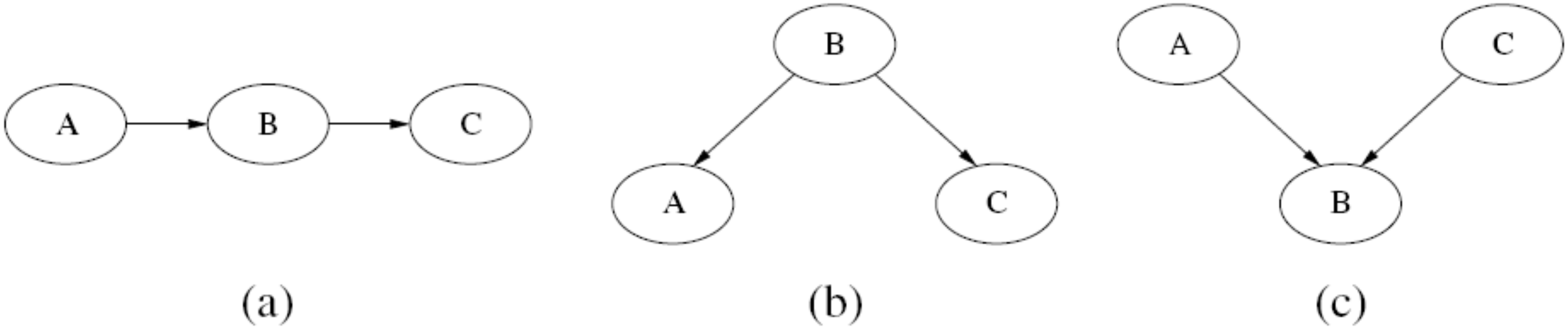


FIGURE 2.4

(a) Causal chain; (b) common cause; (c) common effect.

$$P(C|A \wedge B) = P(C|B) \equiv A \perp\!\!\!\perp C | B$$

MARKOV PROPERTY

- **Markov property:** there are no direct dependencies in the system being modeled which are not already explicitly shown via arcs.
- **d-separation** *A set of nodes* **E** *d-separates two other sets of nodes* **X** *and* **Y** *if every path from a node in* ***X to a node in Y is blocked given E***
- If **X and Y** are **d-separated by E, then X and Y are conditionally independent** given E

- **d-separation (direction-dependent separation)**
- The conditional independence in $A \perp\!\!\!\perp C \mid B$
- means that knowing the value of B blocks information about C being relevant to A, and viceversa
- *lack of information* about B **blocks** the relevance of C to A whereas learning about B **activates** the relation between C and A

- These concepts apply between sets of nodes
- Given the Markov property, it is possible to determine whether a set of nodes **X is independent of another set Y, given a set of evidence nodes E**
- **Path (Undirected Path)** *A path between two sets of nodes **X and Y** is any sequence of nodes between a member of **X and a member of Y such that** every adjacent pair of nodes is connected by an arc (regardless of direction) and no node appears in the sequence twice*

- Blocked path *A path is blocked, given a set of nodes E, if there is a node Z on the path for which at least one of three conditions holds:*
  - *Z is in E and Z has one arc on the path leading in and one arc out (chain).*
  - *Z is in E and Z has both path arcs leading out (common cause)*
  - *Neither Z nor any descendant of Z is in E, and both path arcs lead in to Z (common effect).*

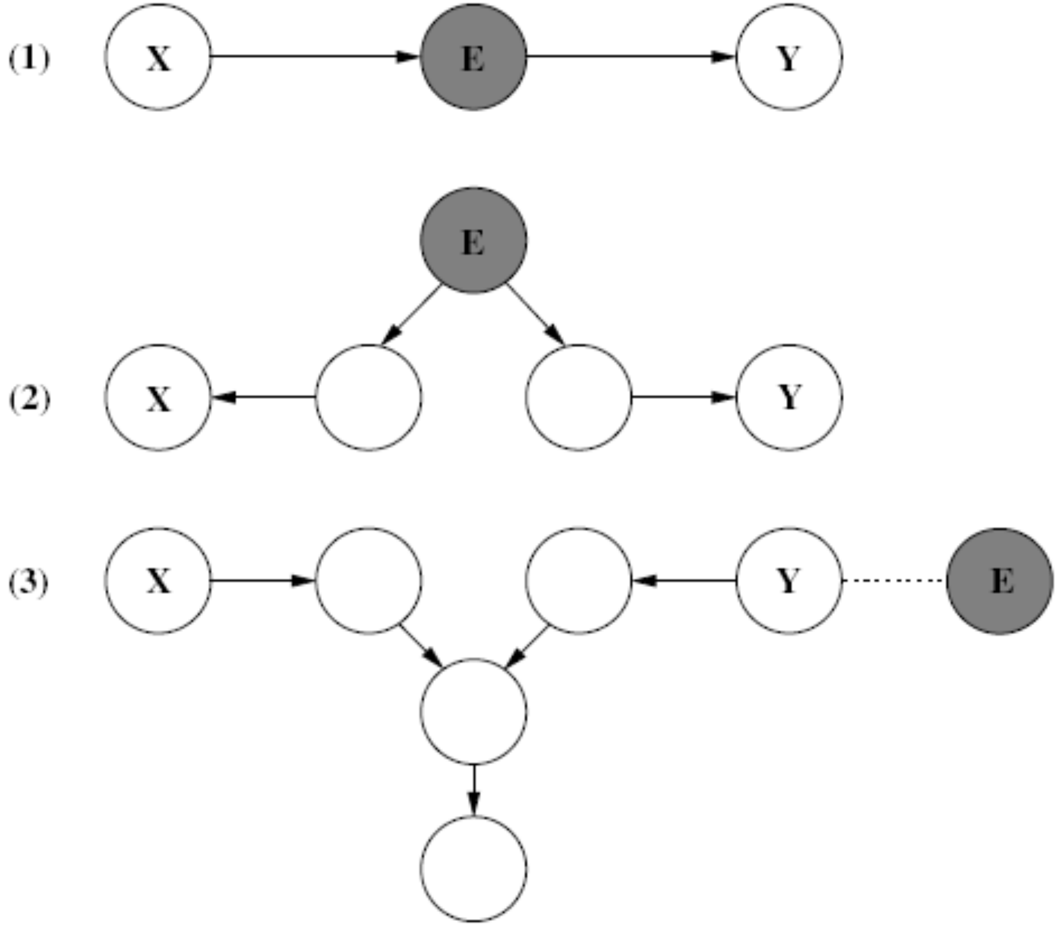


FIGURE 2.5

Examples of the three types of situations in which the path from X to Y can be blocked, given evidence E. In each case, X and Y are **d-separated** by E.

REFERENCIAS

- Kevin B. Korb, Ann E. Nicholson.

  Bayesian Artificial Intelligence, CHAPMAN & HALL/CRC. England, 2004.