



REDES BAYESIANAS

José Manuel Molina López
Grupo de Inteligencia
Artificial Aplicada

CONTENIDO

Introducción

Independencia Condicional

Modelo Grafo y separabilidad

Propagación de la incertidumbre

Modelo simplificado de propagación

ORIGEN

Generalmente es fácil para un experto del dominio especificar qué relaciones de dependencia condicional se dan

- Determinar la topología de la red
- Especificar las probabilidades condicionales de los nodos con dependencia directas
- Calcular cualquier otro valor de probabilidad

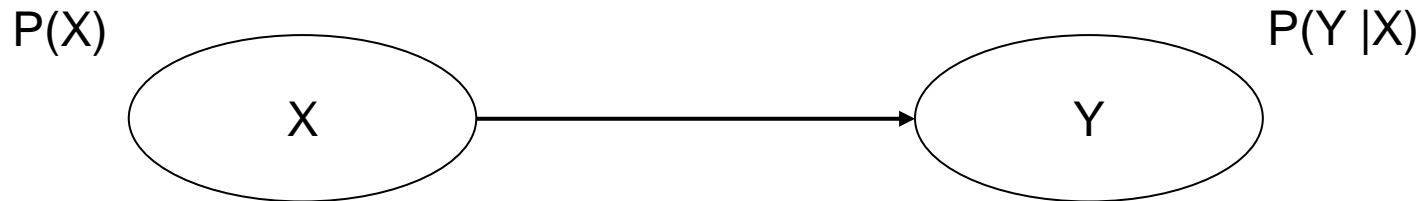
Denominación:

- Redes de creencia
- Redes probabilísticas
- Redes causales
- Mapas de conocimiento

QUÉ SON?

- Estructura de datos para representación de conocimiento incierto
- Representa la dependencia entre variables, y especifica en forma concisa la distribución de probabilidad conjunta
- Representación gráfica
- Cada nodo de la red representa una variable aleatoria
- Un arco del nodo X al nodo Y, significa que la variable X tiene una influencia sobre Y
- Cada nodo X tiene una tabla de probabilidad condicional que cuantifica el efecto que los padres de X tienen sobre X.
- Es un grafo dirigido acíclico (GDA)

PROBABILIDAD CONDICIONAL



X	$P(y X)$	$P(\neg y X)$
V	0.7	0.3
F	0.01	0.99

$$P(y|x) + P(\neg y|x) = 1$$

$$P(y|\neg x) + P(\neg y|\neg x) = 1$$

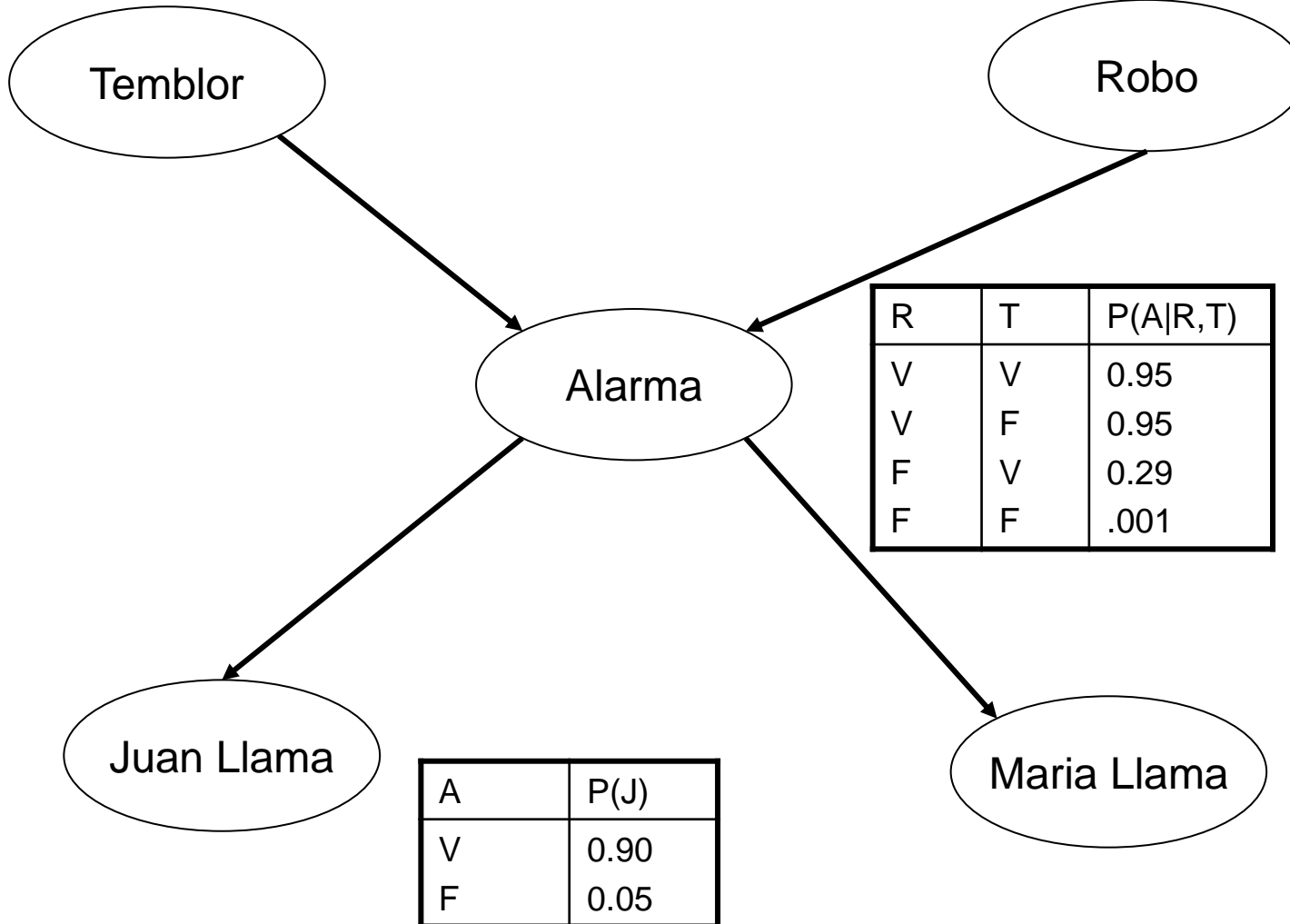
$$P(y) = P(y|x)P(x) + P(y|\neg x)P(\neg x)$$

$$P(\neg y) = P(\neg y|x)P(x) + P(\neg y|\neg x)P(\neg x)$$

EJEMPLO

$$P(T)=0.001$$

$$P(R)=0.002$$



SEMÁNTICA

Dos puntos de vista sobre una RB:

Como representación de la distribución de probabilidad conjunta (DPC)

- Útil para entender cómo construir redes

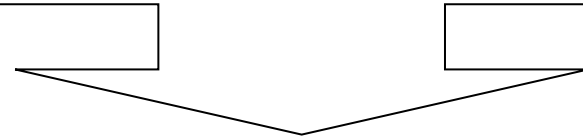
Como representación de un conjunto de aseveraciones de independencia condicional

- Útil para diseñar procedimientos de inferencia

Ambos puntos de vista son equivalentes

INDEPENDENCIA CONDICIONAL

Relación entre causalidad y correlación ?



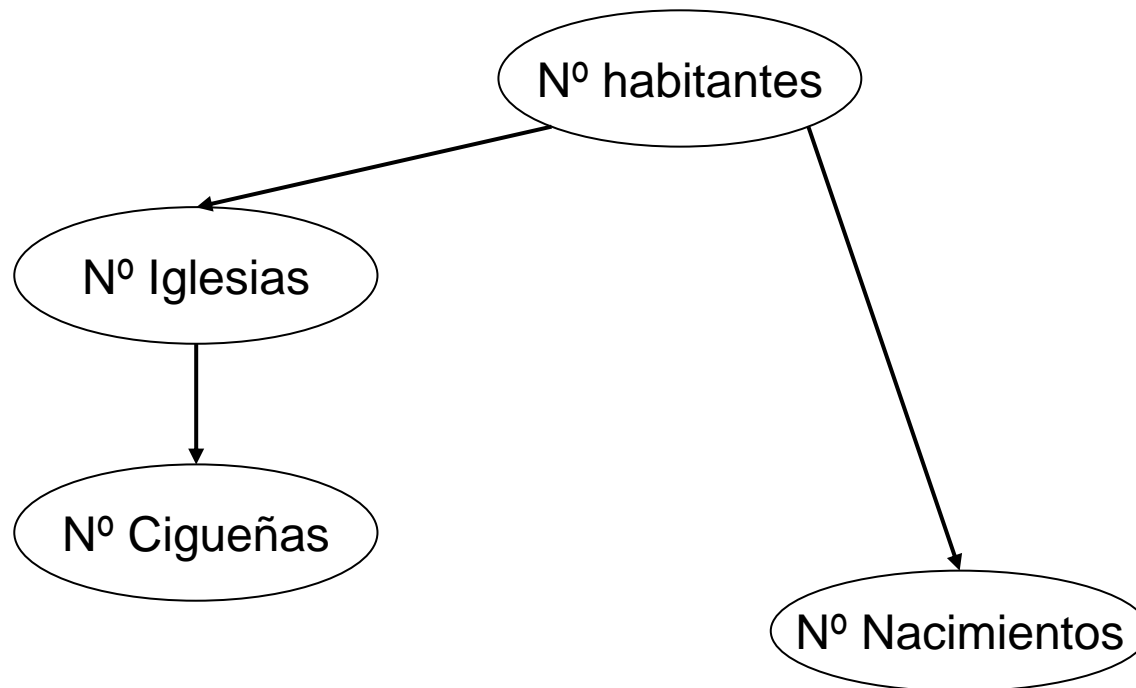
Causalidad \Rightarrow Correlación

Correlación NOT \Rightarrow Causalidad

CORRELACIÓN - CAUSALIDAD

Relación entre causalidad y correlación ?

Ej.: un estudio demostró que hay una fuerte correlación entre el número de cigüeñas de una localidad y el número de nacimientos

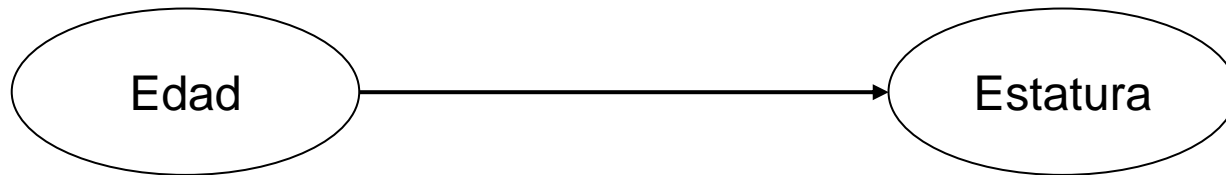


DEPENDENCIAS ENTRE NODOS

Independencia



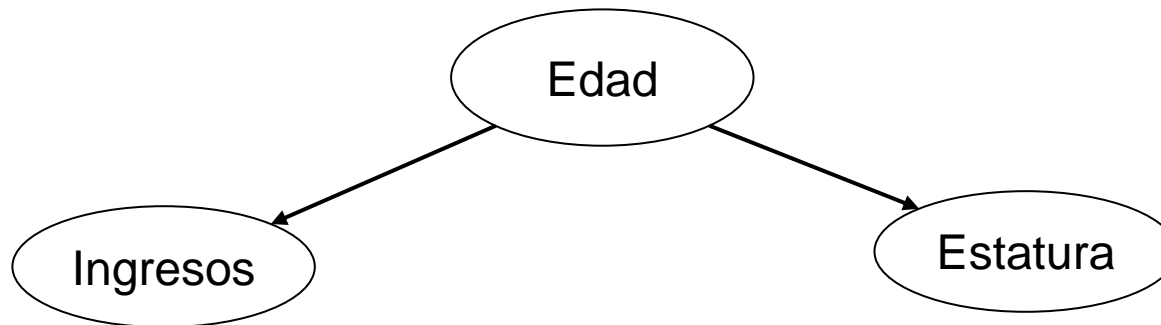
Dependencia causal (tb correlación)



El sentido de la flecha indica Influencia causal

DEPENDENCIAS ENTRE NODOS

Dependencia causal entre un nodo padre y 2 hijos



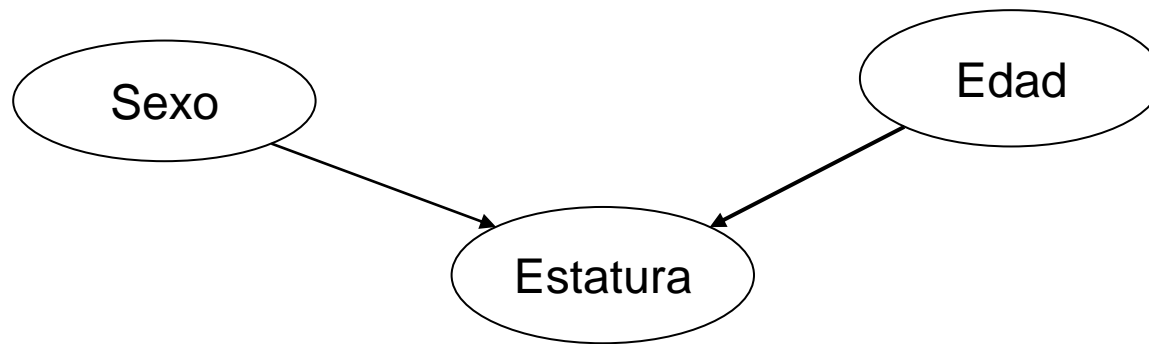
Dependencia causal de 3 nodos en cadena



Independencia condicional

DEPENDENCIAS ENTRE NODOS

Dependencia causal entre 2 padres y 1 nodo hijo



Dependencia condicional

NODOS LINEALES



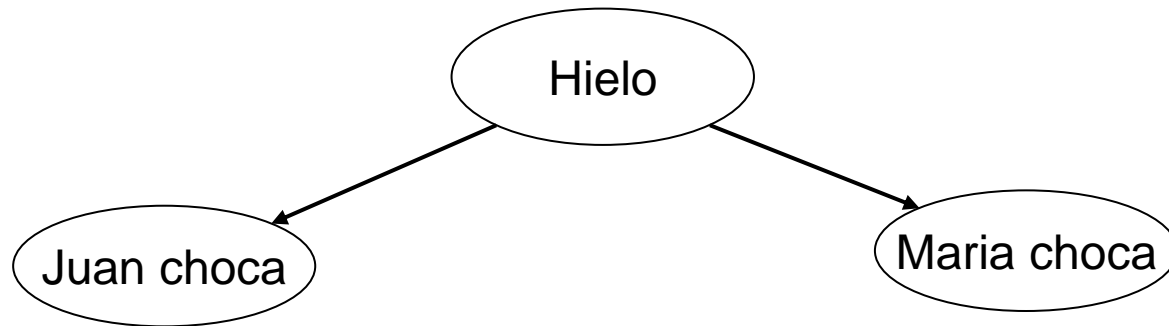
Sin ninguna información adicional L y R son dependientes

- Evidencia 1: T
- R y L son independientes dado T

La evidencia puede ser transmitida a través de un nodo lineal a menos que esté instanciado

En un nodo lineal T los nodos vecinos son condicionalmente independientes respecto a T, es decir, son dependientes si T no está instanciado y viceversa.

NODOS DIVERGENTES



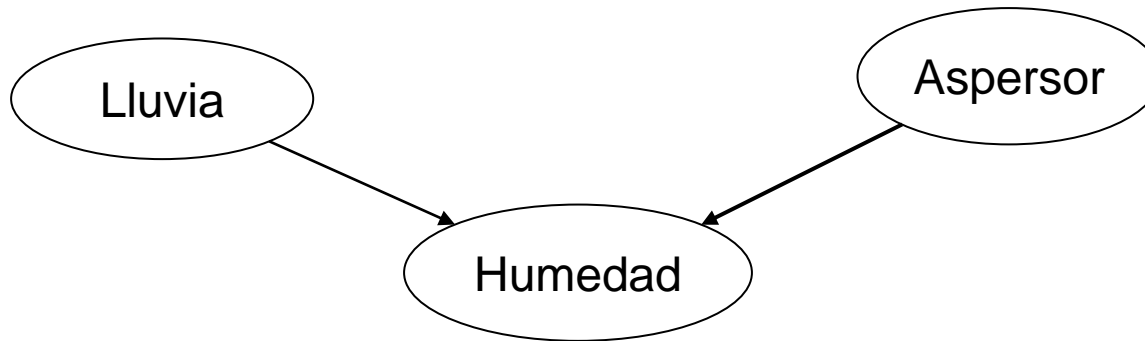
Sin ninguna información adicional J y M son dependientes

- Evidencia 1: $\neg H$
- J y M son independientes dado $\neg H$

La evidencia puede ser transmitida a través de un nodo divergente a menos que esté instanciado

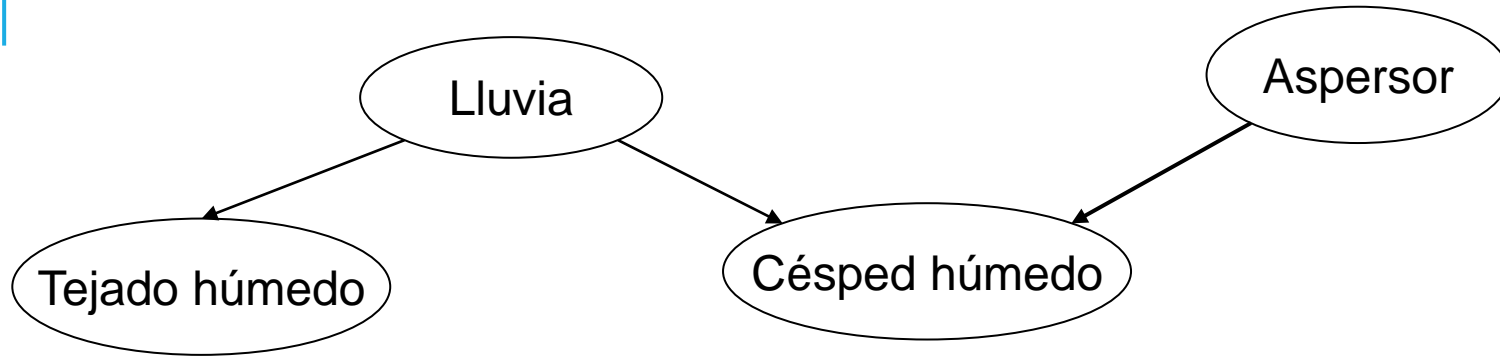
En un nodo divergente H sus hijos son condicionalmente independientes respecto a H, es decir, son dependientes si H no está instanciado y viceversa.

NODOS CONVERGENTES



- ☐ Sin ninguna información adicional L y A son independientes
 - ☐ Evidencia 1: H
 - ☐ L y A son dependientes dado H
- ☐ La evidencia puede ser transmitida a través de un nodo convergente si no está instanciado.
- ☐ En un nodo convergente H no instanciado, sus padres son independientes, pero son condicionalmente dependientes dado H

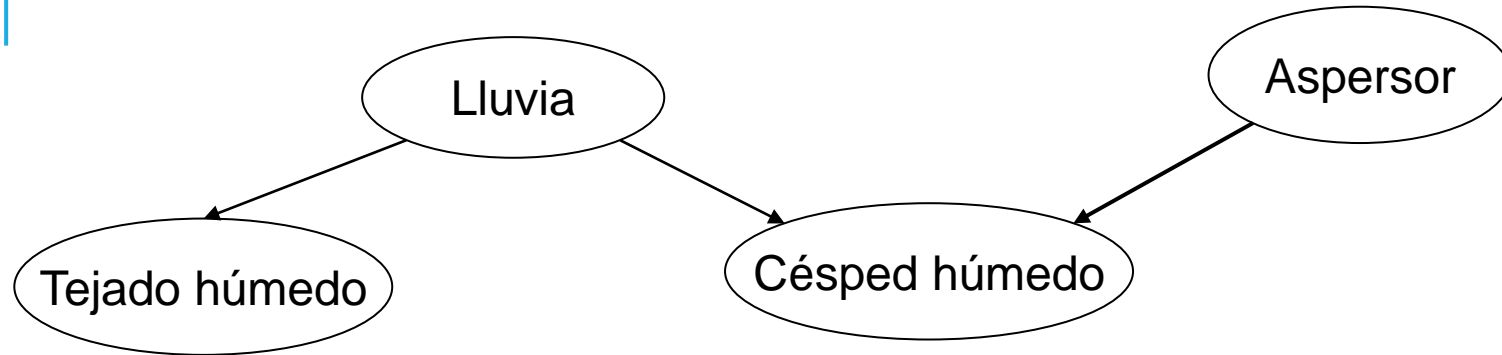
EXPLAINING AWAY



C es un nodo convergente para L y A

L es divergente para T y C

EXPLAINING AWAY



C es un nodo convergente para L y A

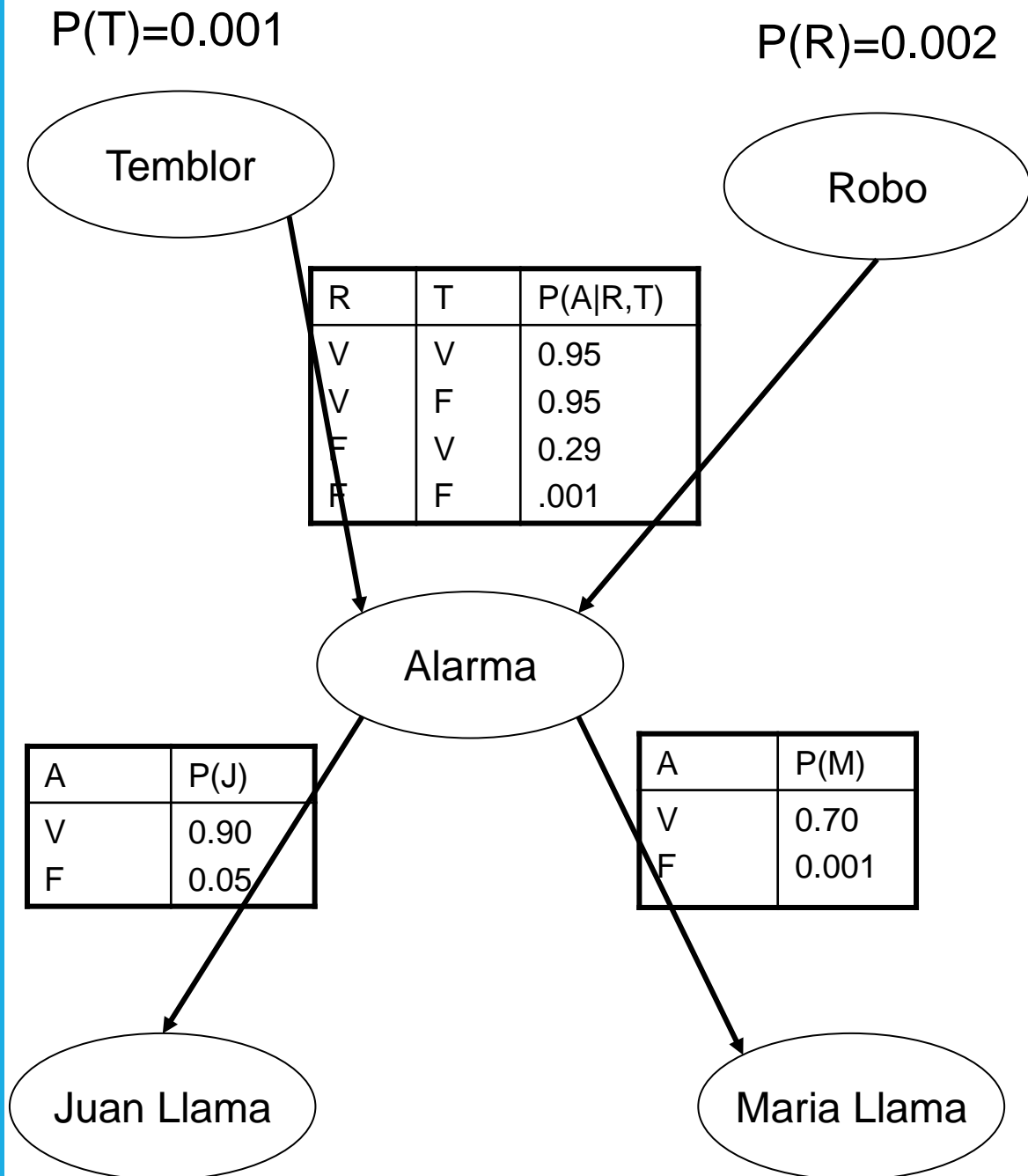
L es divergente para T y C

- Evidencia 1: C
- L y A son dependientes dado C
- Evidencia 2: T
- *Explaining away*: Aceptamos L y descartamos A

RESUMEN DE LAS PROPIEDADES DE INDEPENDENCIA CONDICIONAL

1. Independencia a priori de los nodos que no tienen ningún antepasado común
2. Independencia condicional de los nodos hermanos con respecto a su padre
3. Independencia condicional entre un nodo y los antepasados de sus padres
4. Dependencias condicionales por descendientes comunes instanciados

GRAFOS Y SEPARABILIDAD



GRAFO CONEXOS Y POLIÁRBOLES

Grafo conexo: entre cualquier par de nodos hay al menos un camino (una ruta no dirigida)

- A veces se distingue entre camino abierto y cerrado, que corresponde a ciclos y bucles)

Grafo simplemente conexo o **poliárbol**: entre cualquier par de nodos hay un único camino

Grafo múltiplemente conexo: contiene bucles o ciclos

Árbol: poliárbol en el que cada nodo tiene un solo padre, menos el nodo raíz que no tiene

DEFINICIONES FORMALES

Separación direccional: Dado un GDA conexo (V, E) y una distribución de probabilidad sobre sus variables, P , hay separación direccional si:

- Dado $X \in V$, el conjunto de sus padres, $\text{pa}(X)$, separa condicionalmente a X de cualquier otro conjunto de nodos Y que no tenga descendientes de X , $\text{de}(X)$:

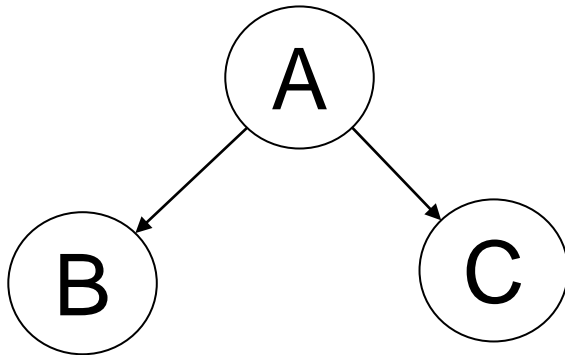
$$P(X \mid \text{padres}(X), Y) = P(X \mid \text{padres}(X))$$

$$X \in V, Y \in V - \{X \cup \text{pa}(X) \cup \text{de}(X)\}$$

DEFINICIONES FORMALES

Red Bayesiana: (V, E, P) GDA más distribución de probabilidad sobre V , que cumple la propiedad de separación condicional

Ejemplo: Sea la siguiente red y $P(a1) = 0.3$; $P(b1/a1) = 0.4$ $P(b1/a2) = 0.2$ $P(c1/a1) = 0.7$ $P(c1/a2) = 0.6$, ¿es una red bayesiana? ¿cumple la independencia condicional?



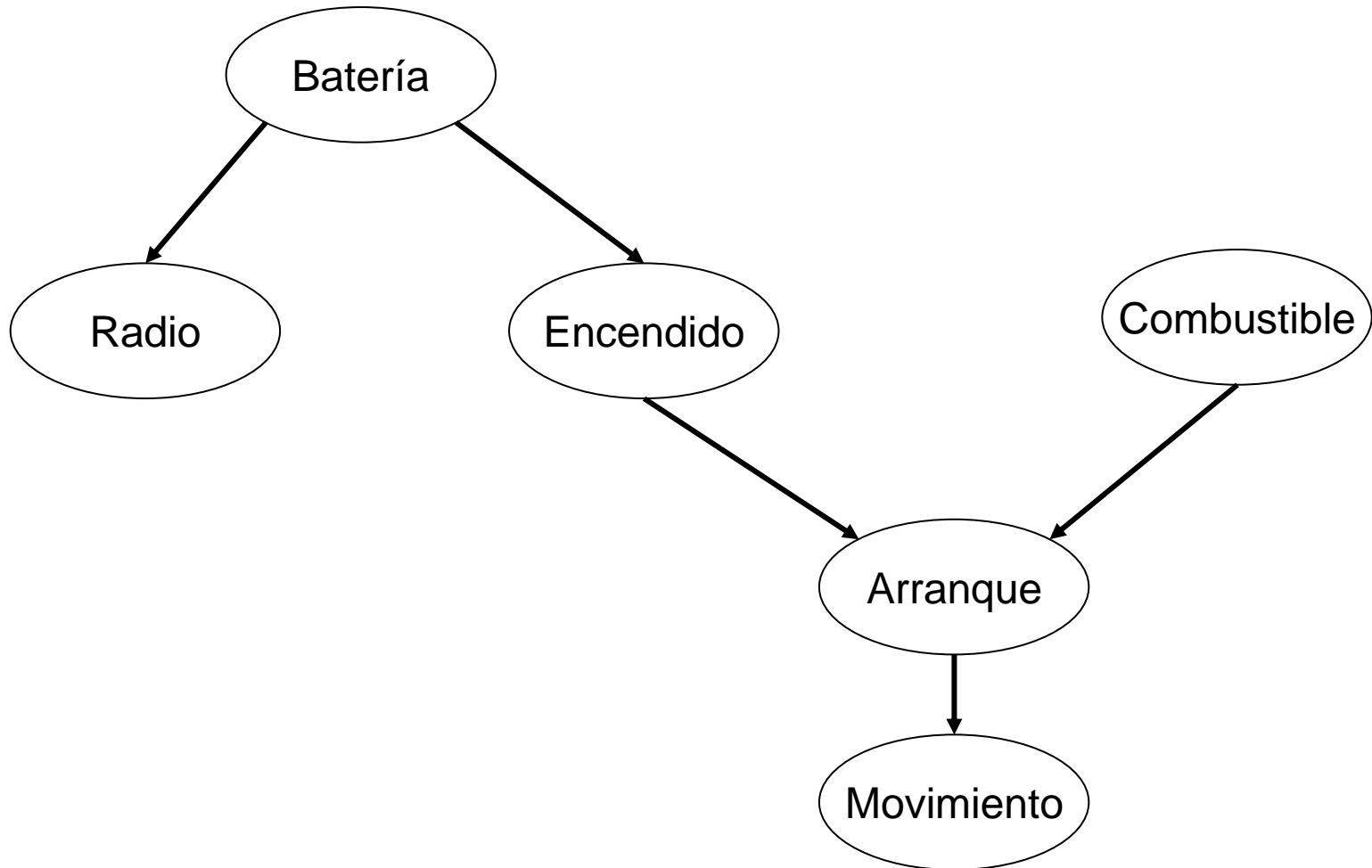
$P(a1, b1, c1) = 0.084$	$P(a1, b1, c2) = 0.036$
$P(a1, b2, c1) = 0.126$	$P(a1, b2, c2) = 0.054$
$P(a2, b1, c1) = 0.084$	$P(a2, b1, c2) = 0.336$
$P(a2, b2, c1) = 0.056$	$P(a2, b2, c2) = 0.224$

INDEPENDENCIA CONDICIONAL

Hemos visto como una RB expresa la independencia entre un nodo y sus antepasados .

¿Es posible saber si un conjunto de nodos X es independiente de otro conjunto Y con base en el conjunto de los nodos de evidencia E ?

EJEMPLO



D- SEPARACIÓN

Un conjunto de nodos E d-separa dos conjuntos de nodos X y Y si cualquier trayectoria no-dirigida de un nodo en X a un nodo en Y es bloqueada en función de E

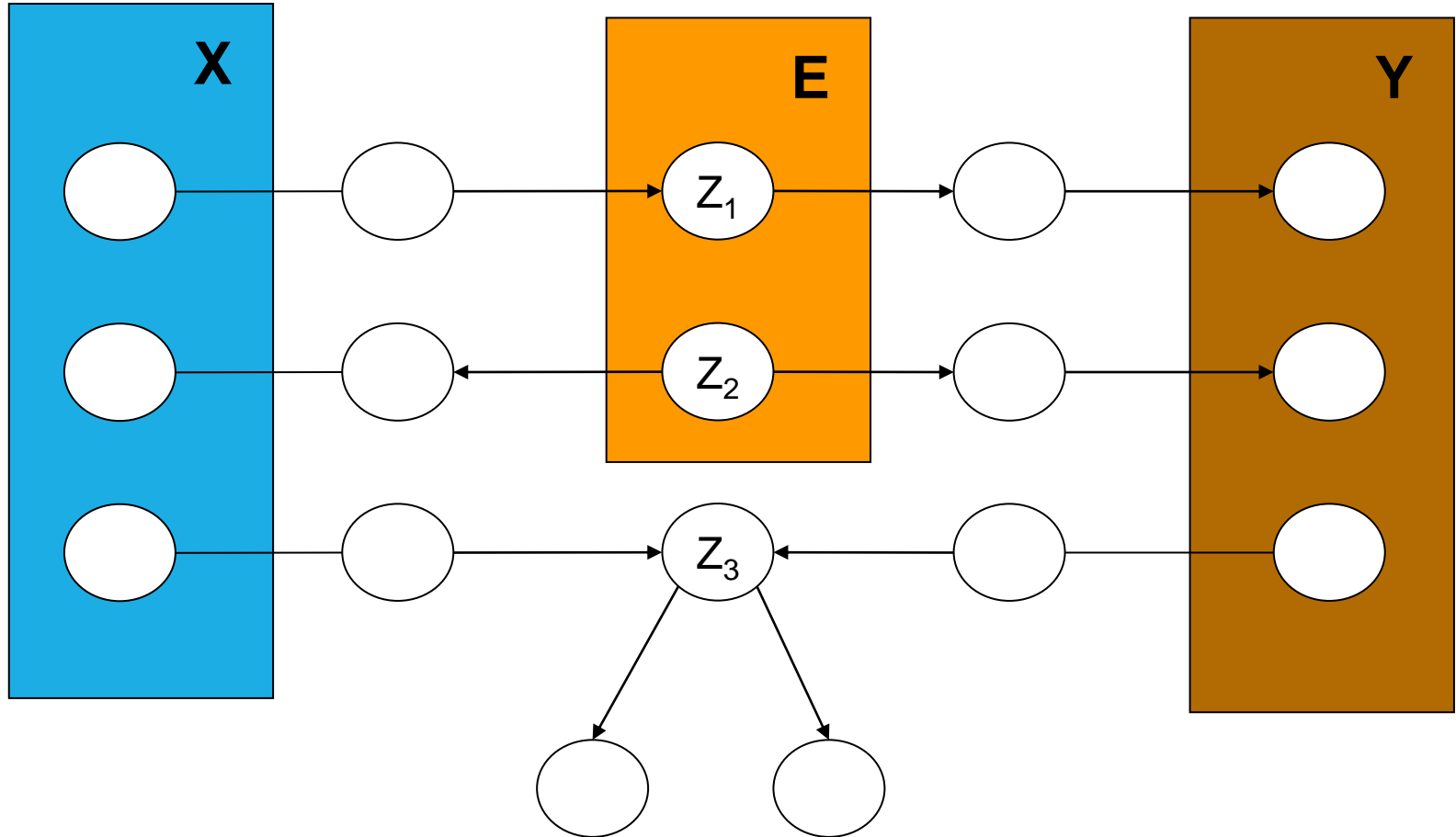
Si la ruta no-dirigida (independiente de la dirección de las flechas) de un nodo X a un nodo Y está d-separada por E , entonces X y Y son condicionalmente independientes dada E

D- SEPARACIÓN

Una ruta es bloqueada en función de un conjunto de nodos E si existe un nodo Z en la ruta que cumple una de las condiciones siguientes:

1. Z está en E, y Z tiene un arco saliente y otro entrante en esa ruta (nodo lineal)
2. Z está en E y Z tiene ambos arcos salientes en esa ruta (nodo divergente)
3. Ni Z ni sus descendientes está en E, y los dos arcos de la ruta son entrantes (nodo convergente)

1



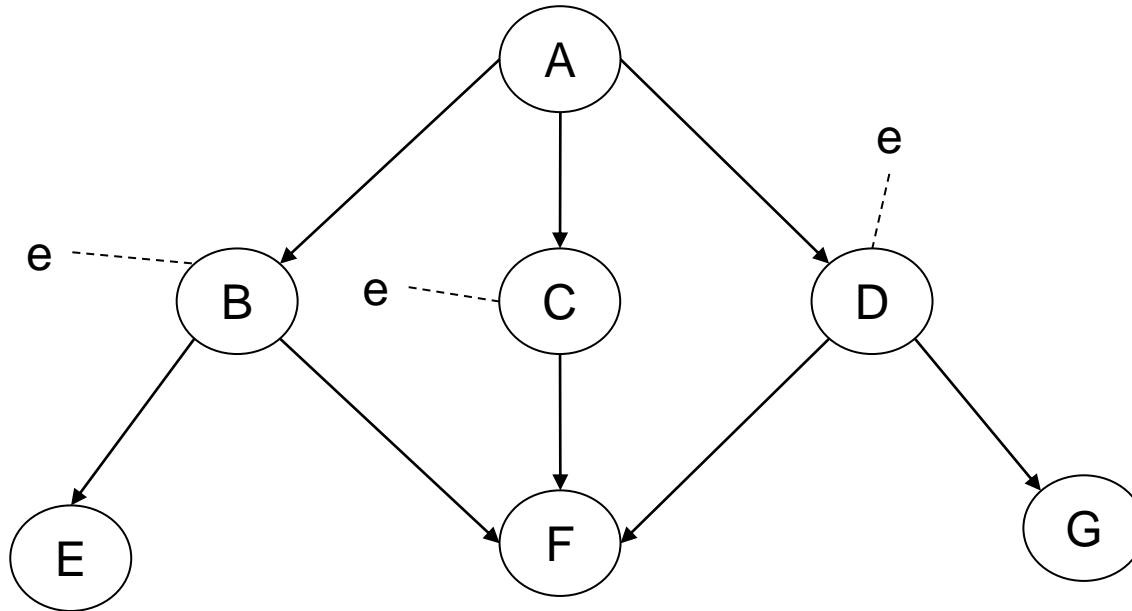
SEPARABILIDAD

Si A y B son d-separadas, entonces cambios en la probabilidad de A no tienen efecto en la probabilidad de B

Si A y B son d-separadas dada la evidencia e, entonces A y B son condicionalmente independientes dado e: $P(A \mid B, e) = P(A \mid e)$

Si A y B no son d-separadas, entonces son d-conectadas

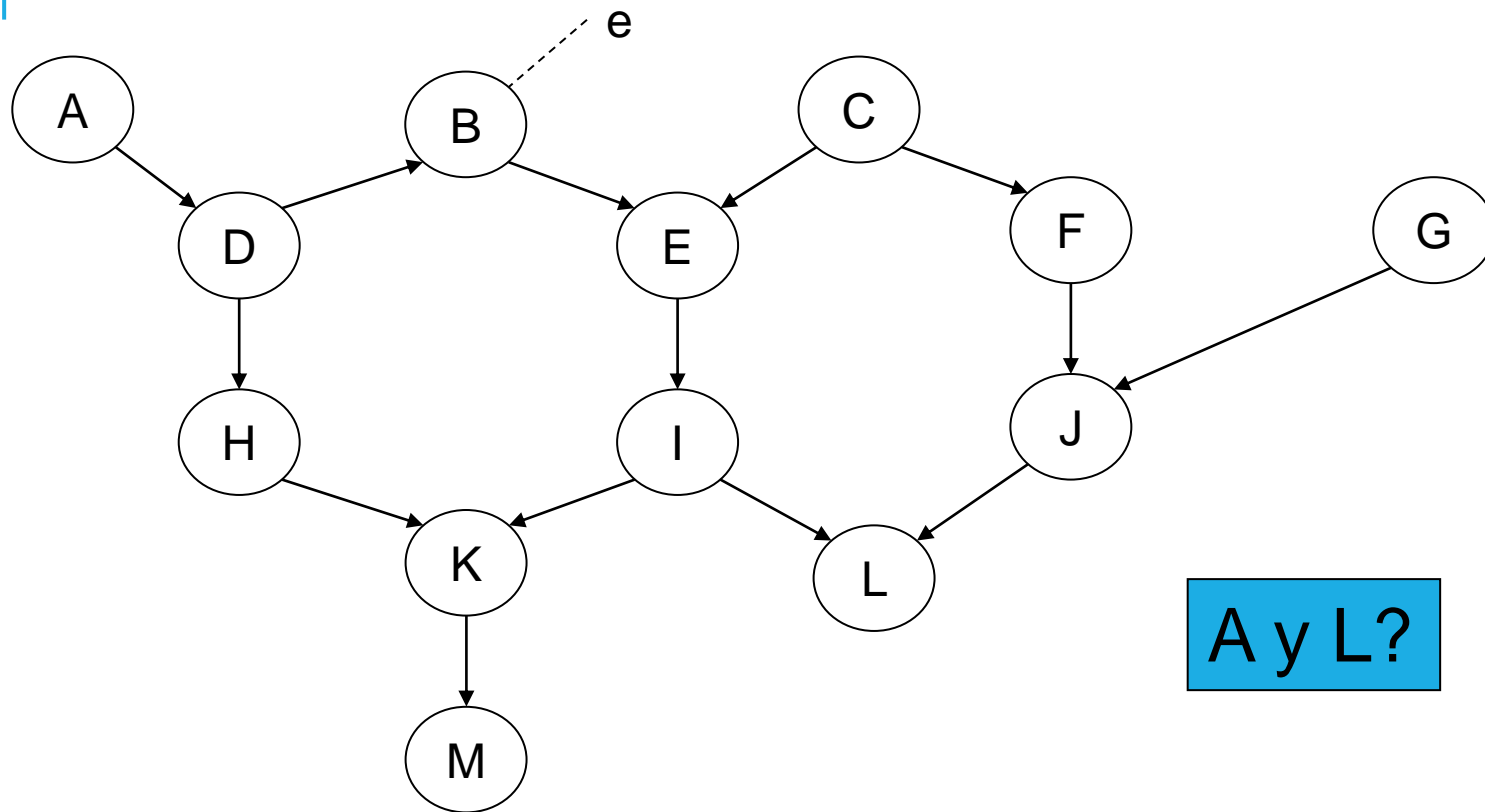
EJEMPLO



A, E, G, F?

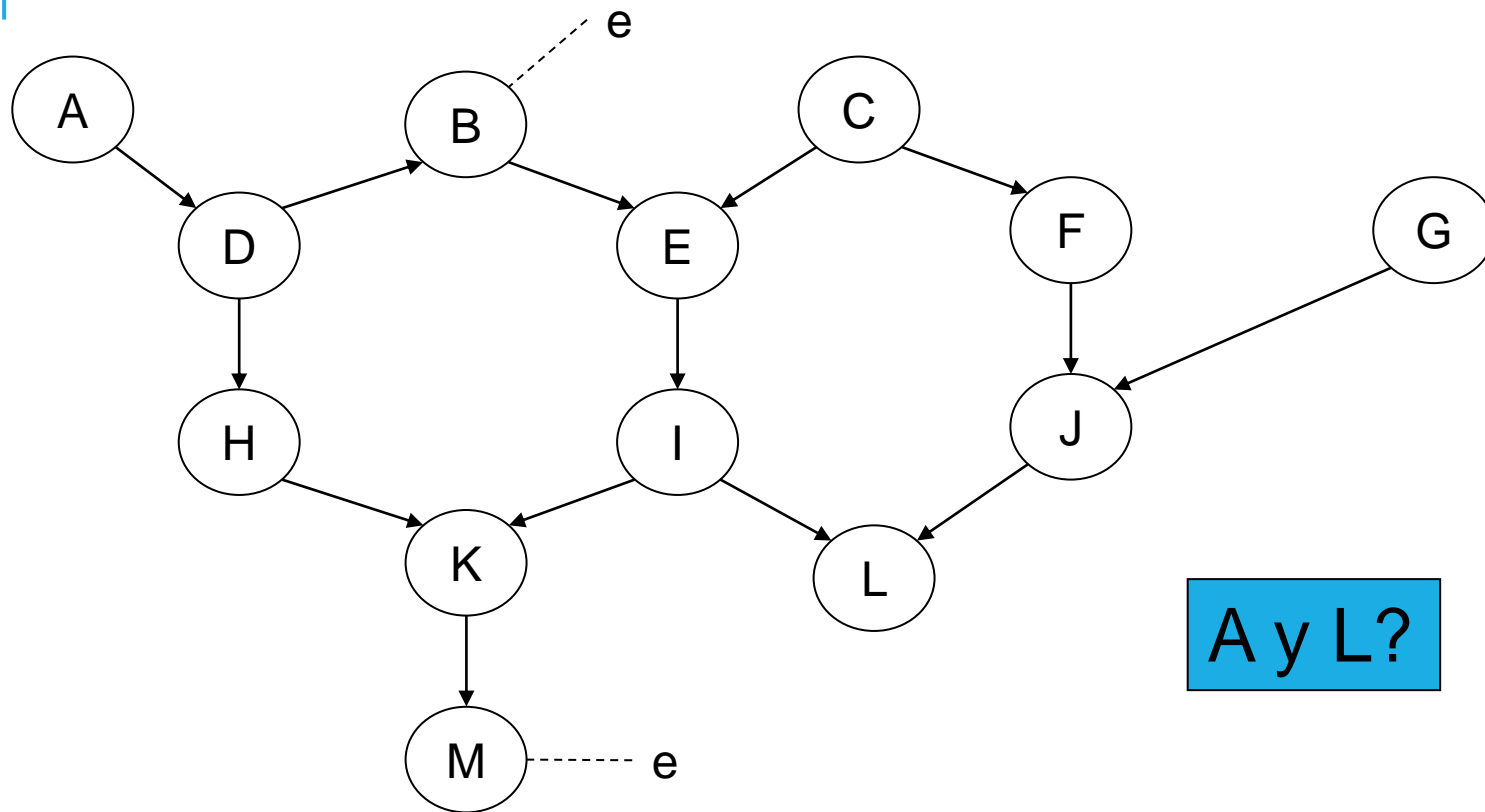
F está d-separada del resto de las variables no-instanciadas A, E y G

EJEMPLO



A y L están d-separadas

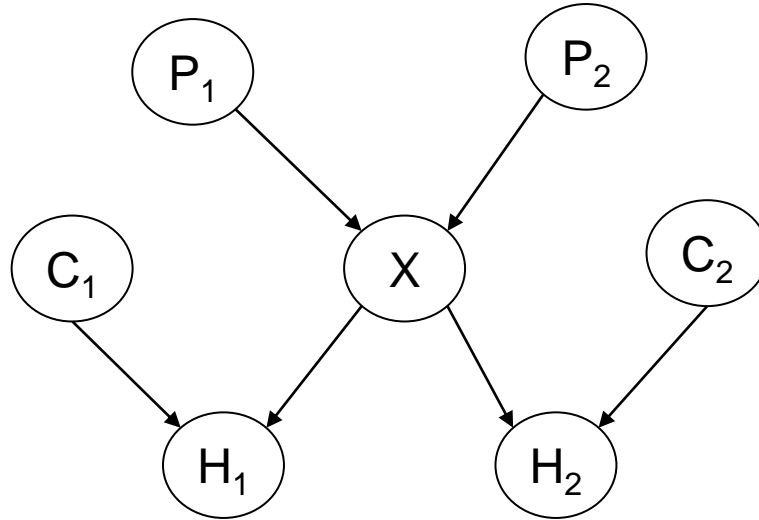
EJEMPLO



A y L?

A y L están d-conectadas

SÁBANA DE MARKOV



□ **Sábana (manto) de Markov** de X : padres de X , hijos de X y cónyuges de X (comparten hijos)

▣ Si se instancian todas, X queda d-separada del resto de la red

PROPIEDADES DE INDEPENDENCIA CONDICIONAL

Una RB es representación correcta del dominio si cada nodo es cond. independiente respecto de antepasados de sus padres

→ Escoger a los padres de manera que se satisfaga la condición anterior

En el ejemplo, no hay relación directa entre el hecho de que Maria o Juan llamen y el que se produzca un terremoto o un robo, relación mediada por el hecho de que suene la alarma

- $P(M \mid J, A, T, R) = P(M \mid A)$
- $P(J \mid J, A, T, R) = P(J \mid A)$

METODOLOGÍA DE CONSTRUCCIÓN

1. Escoger conjunto de variables
2. Definir un orden parcial para el conjunto de variables; primero los nodos causales y luego los nodos efecto
3. Mientras queden variables
 - a) Escoger siguiente variable X_i y añadir nodo a la RB
 - b) Asigne $\text{Padres}(X_i)$ a un conjunto mínimo de nodos presente en la red, de manera que sea satisfecha la propiedad de independencia condicional
 - c) Elaborar la tabla de probabilidad condicional de X_i

Este método garantiza la obtención de redes acíclicas

Evita la redundancia en la definición de probabilidades

Evita que se violen los axiomas de probabilidad

INFERENCIA EN RB

Inferencia o propagación de probabilidades: efectos de la evidencia propagados por la red para saber probabilidades a posteriori

Propagación: dar valores a ciertas variables (evidencia), y obtener la probabilidad posterior de las demás variables

PROBABILIDADES CONJUNTAS

Una RB proporciona una descripción completa del dominio

Cualquier elemento de $P(X_1, \dots, X_n)$ de la DPC se puede calcular a partir de la red

$$P(x_1, \dots, x_n) = \prod_{i=1..n} P(x_i \mid \text{Padres}(x_i))$$

CONSTRUCCIÓN

$$P(x_1, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1}, \dots, x_1)$$

$$P(x_1, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \dots P(x_2 | x_1) P(x_1)$$

$$P(x_1, \dots, x_n) = \prod_{i=1..n} P(x_i | x_{i-1}, \dots, x_1)$$

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{Padres}(X_i)),$$

si $\text{Padres}(X_i) \subseteq \{x_{i-1}, \dots, x_1\}$

Para satisfacer esa condición → Etiquetar los nodos de forma consistente con el orden parcial implícito en la RB.

PROBABILIDADES CONJUNTAS

$$P(x_1, \dots, x_n) = \prod_{i=1..n} P(x_i \mid \text{Padres}(x_i))$$

Ejemplo: Probabilidad de que la Alarma suene, no haya Robo ni Terremoto, y Juan y Maria llamen

$P(A, \neg R, \neg T, J, M)$?

$$P(A \mid \neg R, \neg T) P(\neg R) P(\neg T) P(J \mid A) P(M \mid A) = \\ 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$

COMPACTACIÓN

Una RB es más compacta que la distribución de probabilidad conjunta correspondiente
→ permite manejar muchas evidencias sin el problema del crecimiento exponencial

Sistema localmente estructurado (sparse system).

→ crecimiento lineal (en vez de exponencial)

Si las variables de una RB reciben influencias directas de un promedio de k variables, y hay un total de N variables booleanas, entonces la RB queda especificada por $N \cdot 2^k$ probabilidades

TIPOS DE INFERENCIA

Usando el ejemplo de la alarma

Modelo diagnóstico: efectos (síntomas) \rightarrow causas (diagnóstico)

- $P(R | J), P(R | J, M)$

Modelo causal: Causas \rightarrow efectos

- $P(J | R), P(M, R)$

Inferencias intercausales: entre las causas de un efecto común

- $P(R | A, T)$

Inferencias mixtas: combinación de las anteriores

- $P(A | J, \neg T), P(R | J, \neg T)$

TIPOS DE INFERENCIA

Además de estimar la probabilidad de cierto eventos (la variable de consulta), las RB permiten:

- Estimar que variables de evidencia hay que observar para obtener información útil
- Hacer análisis de sensibilidad: determinar que variables tienen más peso en las probabilidades de la variables consultadas
- Explicar al usuario los resultados de una inferencia probabilista

EVIDENCIA TOTAL VS PARCIAL

Evidencia dura (hard).

Conocimiento determinista:
 $P(A)=1$ ó $P(A)=0$. Al asignar
evidencia dura al nodo se le
llama **instanciación**

Evidencia parcial (soft).

Conocimiento probabilístico
(distinto a 0 y a 1). Incluye a
las probabilidades a priori y
a las actualizadas tras
instanciarse alguna variable

INFERENCIA EN RB

Método más general:

$$P(X_i|E) = \frac{P(X_i, E)}{P(E)} = \frac{\sum_{Z=\{X_j\}, X_j \notin \{E \cup X_i\}} P(X_i, Z, E)}{\sum_{Z=\{X_j\}, X_j \notin \{E\}} P(Z, E)} = \alpha \sum_{Z=\{X_j\}, X_j \notin \{E \cup X_i\}} P(X_i, Z, E)$$

Método no eficiente (N-p completo)

RB almacena de forma eficiente la TPC

¿Inferencia eficiente? Sólo en casos particulares: árboles y extensiones

INFERENCIA EN RB

Inferencia exacta por enumeración – Ejemplo

$P(\text{Robo}/\text{Juanllama} = \text{true}, \text{Mariallama} = \text{true})????$

X: variable pregunta: **Robo (R)**

E: cjto variables evidencia, E_1, \dots, E_m : **Juanllama (J), Mariallama(M)**

Y: variables ocultas, Y_1, \dots, Y_l : **Alarma (A), Terremoto (T)**

Para hacer inferencia se aplica la ecuación:

$$P(X/e) = \alpha P(X, e) = \alpha \sum P(X, \textcolor{red}{y}, e)$$

$$P(R/j,m) = \alpha P(R, j,m) = \alpha \sum P(R, \textcolor{red}{t}, \textcolor{red}{a}, j,m)$$

INFERENCIA EN RB

Inferencia exacta por eliminación de variables – Ejemplo

$$P(R/j,m) = \alpha P(R, j,m) = \alpha \sum P(R, t, a, j,m)$$

$$\alpha P(r) \sum P(t) \sum [P(a/r, t)P(j/a)P(m/a) =$$

$$= \alpha P(r)[$$

$$P(t) [P(a/r, t)P(j/a)P(m/a) + P(-a/r, t)P(j/-a)P(m/-a)] +$$

$$P(-t) [P(a/r, -t)P(j/a)P(m/a) + P(-a/r, -t)P(j/-a)P(m/-a)]]]$$

- *No repetir cálculos*
- *Evaluar las expresiones de derecha a izquierda*
- *Guardar resultados intermedios*
- *Sumar sobre una variable sólo las partes que dependen de la variable*
- *Algunas variables son irrelevantes para resolver la pregunta!*

INFERENCIA APROXIMADA: MUESTREO DIRECTO

No se realiza el cálculo, se simula la red

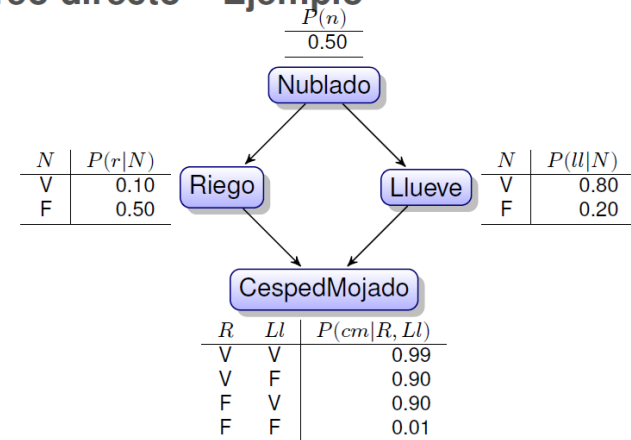
Se muestrea cada variable en orden topológico

La distribución de probabilidad de la que se muestrea es la condicionada a los valores asignados a los padres

$$P(X_i / \text{Padres}(X_i))$$

En el límite (si el proceso se repite muchas veces) tenemos la distribución de probabilidad conjunta a priori

Muestreo directo – Ejemplo



- Muestreo de $P(\text{Nublado}) = (0.5, 0.5)$: sale **true**
- Muestreo de $P(\text{Riego} / \text{Nublado} = \text{true}) = (0.1, 0.9)$: sale **false**
- Muestreo de $P(\text{Lluvia} / \text{Nublado} = \text{true}) = (0.8, 0.2)$: sale **true**
- Muestreo de $P(\text{CéspedMojado} / \text{Riego} = \text{false}, \text{Lluvia} = \text{true}) = (0.9, 0.1)$: sale **true**

INFERENCIA EN ÁRBOLES

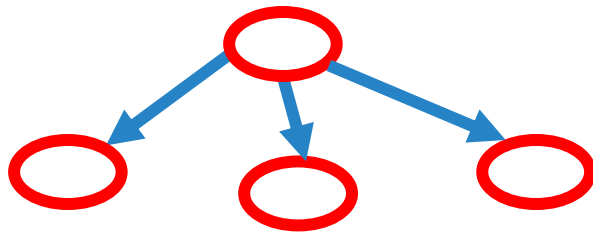
Inferencia o propagación de probabilidades: efectos de la evidencia propagados por la red para saber probabilidades a posteriori

Propagación: dar valores a ciertas variables (evidencia), y obtener la probabilidad posterior de las demás variables

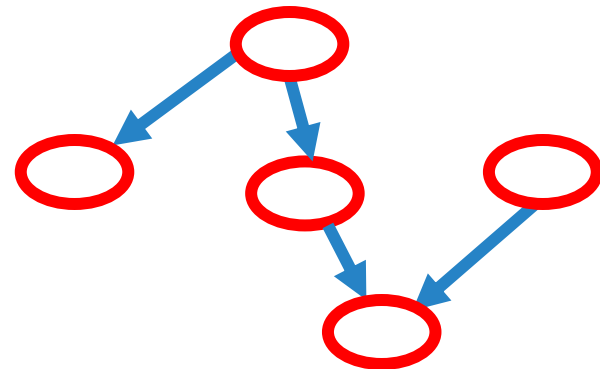
TIPOS DE ESTRUCTURAS

- Redes conectadas en forma sencilla:

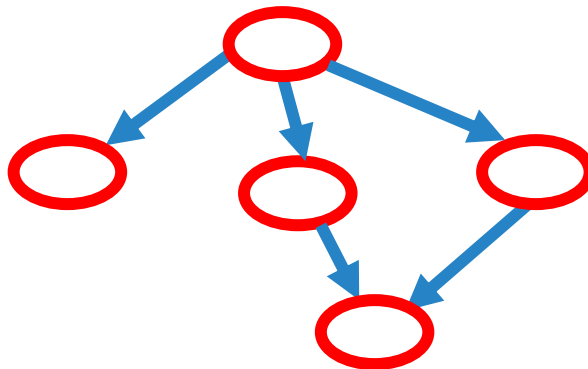
Árboles



Poliárboles

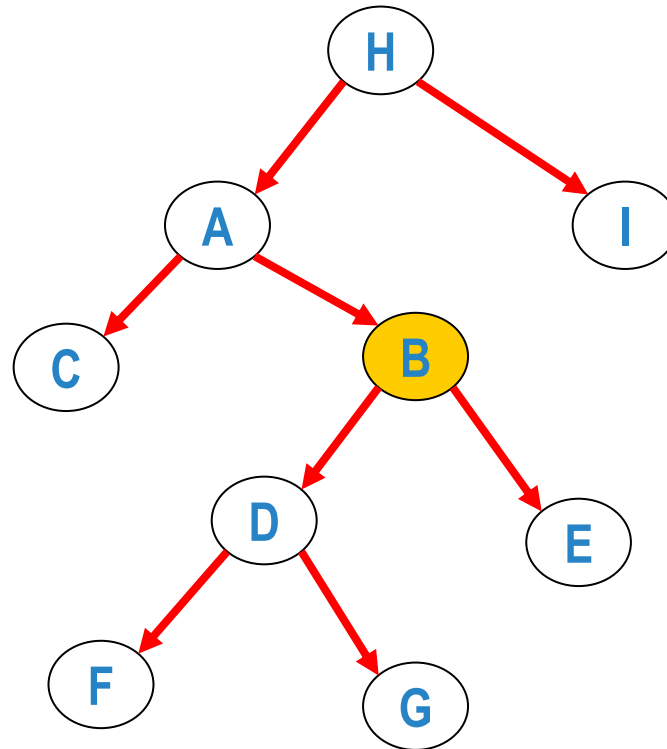


- Redes multiconectadas:



PROPAGACIÓN EN POLIÁRBOLES

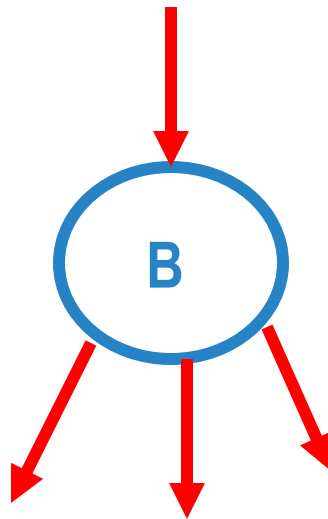
- Cada nodo corresponde a una variable discreta, $B\{B_1, B_2, \dots, B_n\}$ con su respectiva matriz de probabilidad condicional, $P(B|A)=P(B_j|A_i)$:



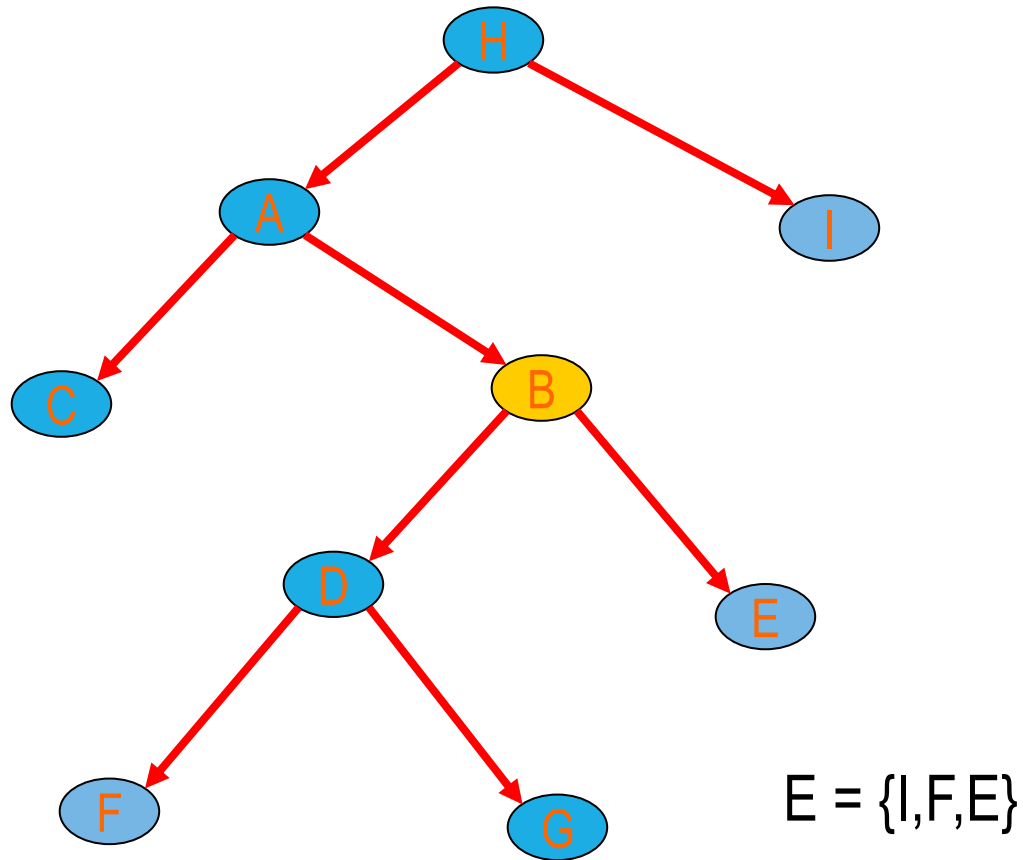
PROPAGACIÓN EN POLIÁRBOLES

- Dada cierta evidencia E (instanciación de ciertas variables) la probabilidad posterior de cualquier variable B, por el teorema de Bayes:

$$P(B_i | E) = P(B_i) P(E | B_i) / P(E)$$



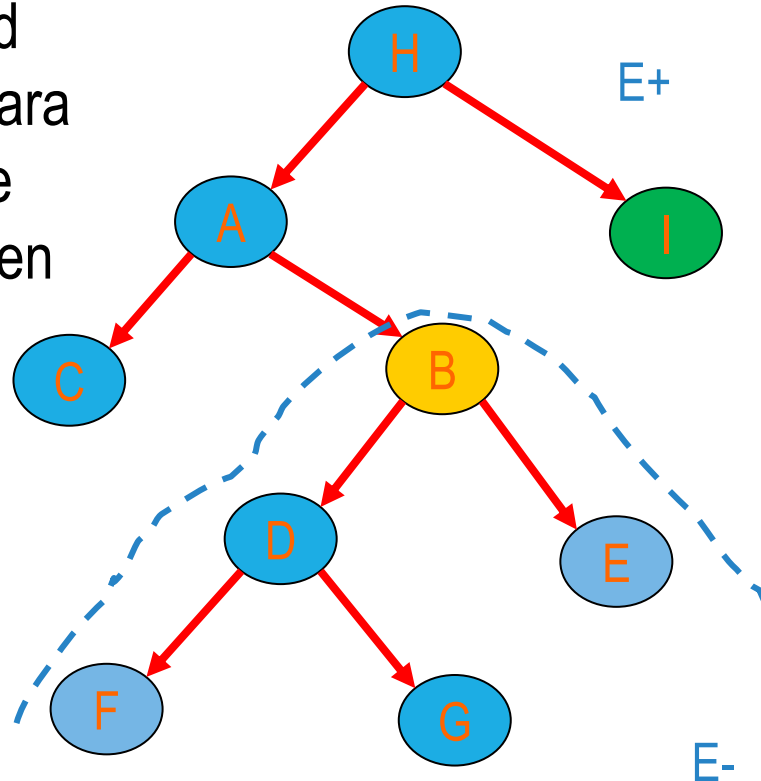
PROPAGACIÓN EN POLIÁRBOLES



PROPAGACIÓN EN POLIÁRBOLES

- Ya que la estructura de la red es un árbol, el Nodo B la separa en dos subárboles, por lo que podemos dividir la evidencia en dos grupos:

- 1) **E-**: Datos en el árbol que cuya raíz es B.
- 2) **E+**: Datos en el resto del árbol



PROPAGACIÓN EN POLIÁRBOLES

Entonces:

$$P(B_i | E) = P (B_i) P (E^-, E^+ | B_i) / P(E)$$

Pero dado que ambos son independientes y aplicando nuevamente Bayes:

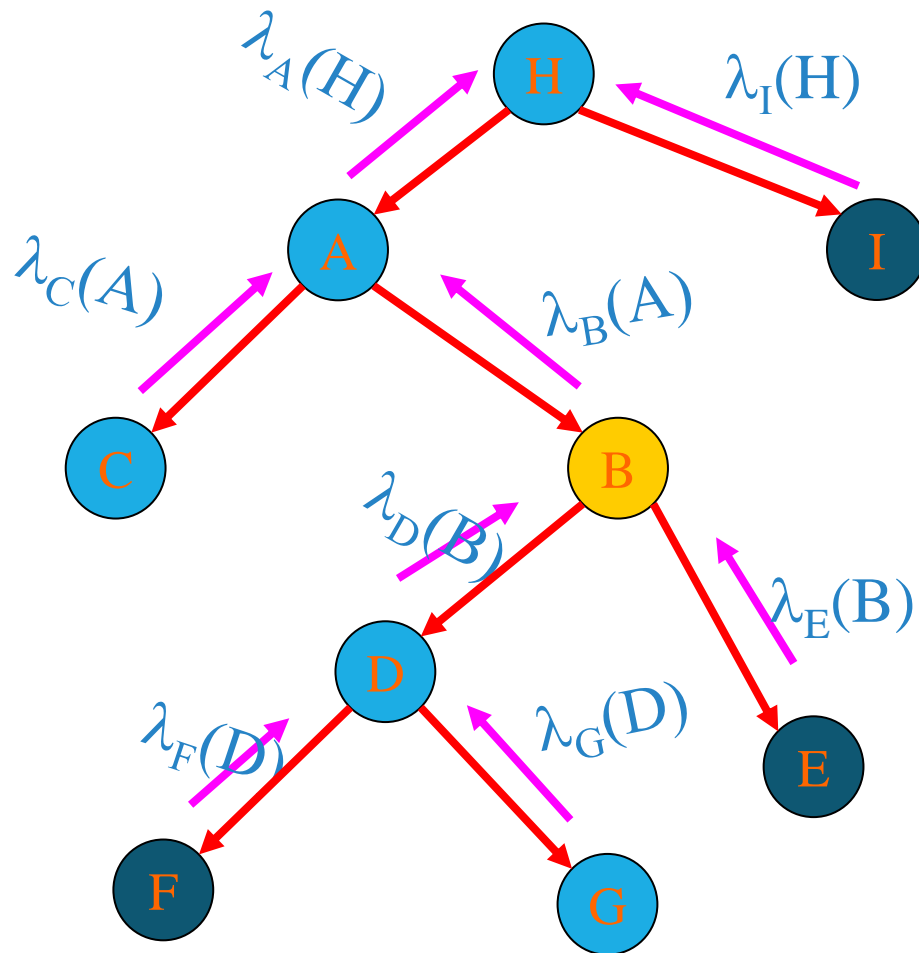
$$P(B_i | E) = \alpha P (B_i | E^+) P(E^- | B_i) = \alpha \lambda(B_i) \pi(B_i)$$

Donde α es una constante de normalización

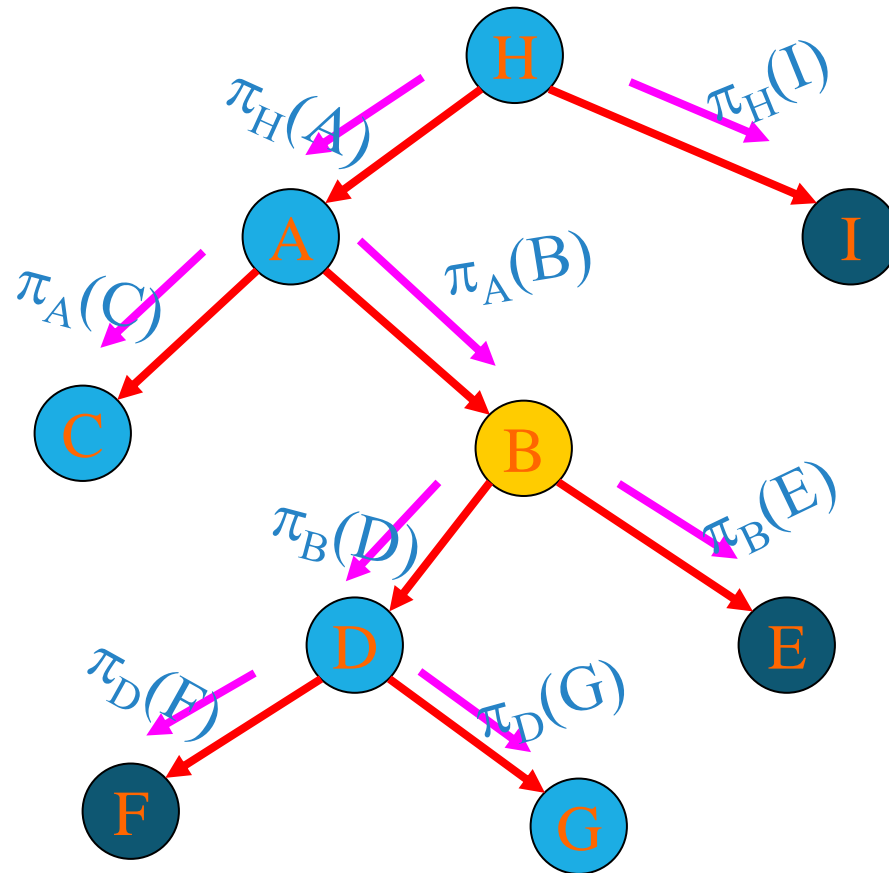
PROPAGACIÓN EN POLIÁRBOLES

- Al instanciarse ciertos nodos, éstos envían mensajes a sus padres e hijos, y se propagan hasta a llegar a la raíz u hojas, o hasta encontrar un nodo instanciado
- Así que la propagación se hace en un solo paso en un tiempo proporcional al diámetro de la red

PROPAGACIÓN Λ



PROPAGACIÓN Π



CONDICIONES INICIALES

- Nodos hoja no conocidos:

$$\lambda (B_i) = [1, 1, \dots]$$

- Nodos asignados (conocidos):

$$\lambda (B_i) = [0, 0, \dots, 1, 0, \dots, 0] \text{ (1 para valor asignado)}$$

$$\pi (B_i) = [0, 0, \dots, 1, 0, \dots, 0] \text{ (1 para valor asignado)}$$

- Nodo raíz:

$$\pi (A) = P(A), \text{ (probabilidad marginal inicial)}$$

CÁLCULO DE Λ -MENSAJES

- Si B es un hijo de A, B tiene k valores posibles y A tiene m valores posibles, entonces $j=1,2,\dots,m$, el λ -mensaje de B a A viene dado por:

$$\lambda_B(a_j) = \sum_{i=1}^k P(b_i / a_j) \lambda(b_i)$$



Ecuación 1

CÁLCULO DE Π -MENSAJES

- Si B es hijo de A y A tiene m valores posibles, entonces para $j=1,2,\dots,m$, el π -mensaje de A a B viene dado por:



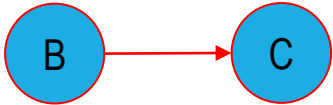
$$\pi_B(a_j) = \begin{cases} \pi(a_j) \prod_{\substack{c \in s(A) \\ c \neq B}} \lambda_c(a_j) & \text{si A no ha sido instanciada} \\ 1 & \text{si } A = a_j \\ 0 & \text{si } A \neq a_j. \end{cases}$$

Ecuación 2

donde $s(A)$ denota al conjunto de hijos de A.

CÁLCULO DE Λ -VALORES

- Si B tiene k valores posibles y $s(B)$ es el conjunto de los nodos C hijos de B, entonces para $i=1,2,\dots,k$, el λ -valor de B viene dado por




A diagram showing a parent node B (blue circle) with a red arrow pointing to a child node C (blue circle).

$$\lambda(b_i) = \begin{cases} \prod_{C \in s(B)} \lambda_C(b_i) & \text{si } B \text{ no ha sido instanciada} \\ 1 & \text{si } B = b_i \\ 0 & \text{si } B \neq b_i. \end{cases}$$

Ecuación 3

CÁLCULO DE Π -VALORES

- Si A es padre de B, B tiene k valores posibles y A tiene m valores posibles, entonces para $i=1,2,\dots,k$, el π -valor de B viene dado por:


$$\pi(b_i) = \sum_{j=1}^m P(b_i / a_j) \pi_B(a_j)$$

Ecuación 4

CÁLCULO DE PROBABILIDAD A POSTERIORI

- Si B es una variable con k posibles valores, entonces para $i = 1, 2, \dots, k$ la probabilidad a posteriori basada en las variables instanciadas se calcula como:



$$P^*(b_i) = \alpha \lambda(b_i) \pi(b_i)$$

Ecuación 5

ALGORITMO. FASE 1- INICIALIZACIÓN

- A. Inicializar todos los λ -mensajes y λ -valores a 1.
- B. Si la raíz A tiene m posibles valores, entonces para $j = 1, \dots, m$, sea:

$$\pi(a_j) = P(a_j).$$

- C. Para todos los hijos B de la raíz A, hacer
 Enviar un nuevo π -mensaje a B usando la ecuación 2.
 (En ese momento comenzará un flujo de propagación debido al procedimiento de actualización C)

ALGORITMO. FASE 2- ACTUALIZACIÓN

Cuando una variable se instancia o una variable recibe un λ o π -mensaje, se usa uno de los siguientes procedimientos de actualización: (A, B ó C)

Procedimiento A

A. Si una variable B se instancia a un valor b_j , entonces

BEGIN

A.1. Inicializar $P^*(b_j) = 1$ y $P^*(b_i) = 0$, para todo $i \neq j$.

A.2. Calcular $\lambda(B)$ usando la ecuación 3.

A.3. Enviar un nuevo λ -mensaje al padre de B usando la ecuación 1.

A.4. Enviar nuevos π -mensajes a los hijos de B usando la ecuación 2.

END

ALGORITMO. FASE 2- ACTUALIZACIÓN

Procedimiento B

B. Si una variable B recibe un nuevo λ -mensaje de uno de sus hijos y la variable B no ha sido instanciada todavía, entonces,

BEGIN

B.1. Calcular el nuevo valor de $\lambda(B)$ usando la ecuación 3.

B.2. Calcular el nuevo valor de $P^*(B)$ usando la ecuación 5.

B.3. Enviar un nuevo λ -mensaje al padre de B usando la ecuación 1.

B.4. Enviar nuevos π -mensajes a los otros hijos de B usando la ecuación 2.

END.

Procedimiento C

C. Si una variable B recibe un nuevo π -mensaje de su padre y la variable B no ha sido instanciada todavía, entonces,

BEGIN

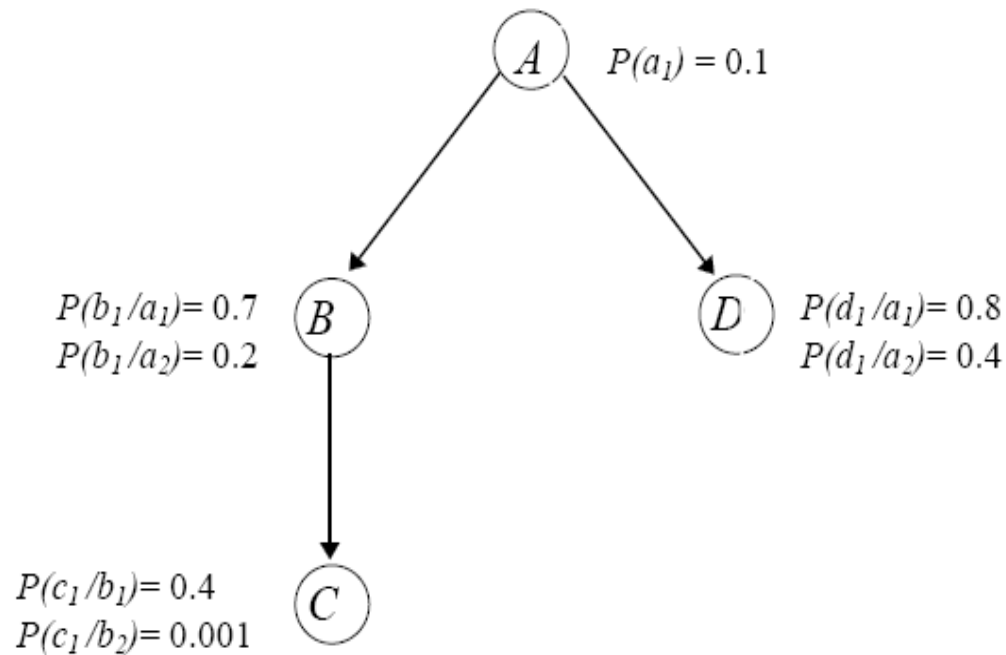
C.1. Calcular el nuevo valor de $\pi(B)$ usando la ecuación 4.

C.2. Calcular el nuevo valor de $P^*(B)$ usando la ecuación 5.

C.3. Enviar nuevos π -mensajes a los hijos de B usando la ecuación 2.

END.

EJEMPLO



“Esposo infiel” :

A esposo infiel

B esposo cena con otra mujer

C esposo es vista cenando con otra mujer

D en el domicilio se reciben llamadas telefónicas sospechosas

a_1 afirmación del hecho, a_2 , negación.

EJEMPLO

1. Calculamos las probabilidades iniciales (a priori) de cada variable

A. Ponemos todos los λ -mensajes y λ -valores a 1.

B. Hacemos $\pi(a_j) = P(a_j)$, para $j = 1, 2$. $\pi(A) = (0.1, 0.9)$.

C. A envía un mensaje a su hijo, B ,

$$\pi_B(a_1) = \pi(a_1) \lambda_D(a_1) = 0,1$$

$$\pi_B(a_2) = \pi(a_2) \lambda_D(a_2) = 0,9$$

B toma entonces nuevos π -valores;

$$\pi(b_1) = P(b_1/a_1) \pi_B(a_1) + P(b_1/a_2) \pi_B(a_2) = 0.7 \cdot 0.1 + 0.2 \cdot 0.9 = 0.25$$

$$\pi(b_2) = P(b_2/a_1) \pi_B(a_1) + P(b_2/a_2) \pi_B(a_2) = 0.3 \cdot 0.1 + 0.8 \cdot 0.9 = 0.75$$

Y con ellos y con los λ -valores de B , se obtienen las probabilidades:

$$P(b_1) = \alpha 0.25 \cdot 1 = 0.25.$$

$$P(b_2) = \alpha 0.75 \cdot 1 = 0.75.$$

EJEMPLO

Ahora, C recibe un π -mensaje por ser hijo de B:

$$\pi_C(b_1) = \pi(b_1) = 0,25$$

$$\pi_C(b_2) = \pi(b_2) = 0,75$$

Y actualiza su π -valor:

$$\pi(c_1) = P(c_1/b_1) \pi_C(b_1) + P(c_1/b_2) \pi_C(b_2) = 0.4 \cdot 0.25 + 0.001 \cdot 0.75 = 0.10075$$

$$\pi(c_2) = P(c_2/b_1) \pi_C(b_1) + P(c_2/b_2) \pi_C(b_2) = 0.89925.$$

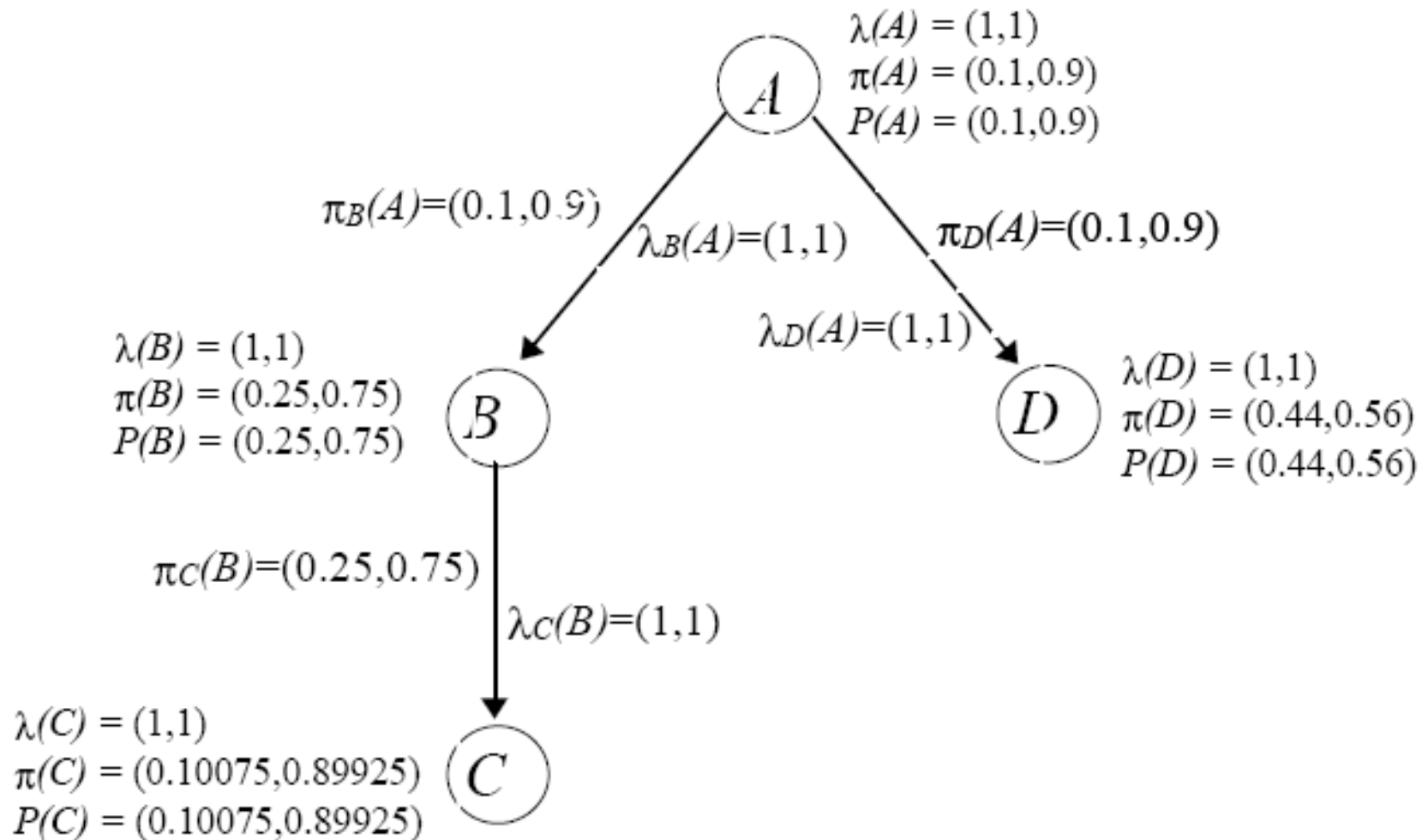
A partir de ellos, calculamos las probabilidades de C, multiplicando por los λ -valores y normalizando:

$$P(c_1) = 0.10075$$

$$P(c_2) = 0.89925.$$

El mismo procedimiento se repite para D, y obtenemos el estado inicial S_0 de la red causal:

EJEMPLO



EJEMPLO

2. Supongamos ahora que nos informan de que el esposo ha cenado con otra mujer, es decir, conocemos ahora con certeza que $B = b_1$.

Esta información se irá transmitiendo por la red, haciendo que las probabilidades a priori de los nodos, $P(X)$ cambien a las probabilidades a posteriori, $P^*(X) = P(X/B = b_1)$. En este caso, al ser la evidencia aportada a favor de la hipótesis que queremos probar, lo lógico será que todas estas probabilidades aumenten. En el momento que una variable se actualiza, comienza un flujo de propagación por la red, que en este caso es el siguiente:

- B informa a su padre mediante un λ -mensaje.
- B informa a su hijo mediante un π -mensaje.
- A su vez, A va a informar a su hijo, D, mediante un π -mensaje.

EJEMPLO

3. Supongamos ahora que tenemos la información de que no se han recibido llamadas telefónicas extrañas en el domicilio, es decir, que sabemos que D ha tomado el valor d_2 .

Nuevamente se iniciará el algoritmo que propagará esta información por la red:

- D enviará un λ -mensaje a su padre, A ,
- A enviará un π -mensaje a su hijo, B .

Pero ahora, al estar B inicializada, el algoritmo se parará ahí, puesto que $P(B) = (1, 0)$, y no podemos permitir que nada cambie ya estos valores. Así, en la ejecución del algoritmo, las variables que ya han sido inicializadas son extremos muertos, donde la propagación se para (en el caso de la propagación en árboles).