Métodos de Inteligencia Artificial

L. Enrique Sucar (INAOE)
esucar@inaoep.mx
ccc.inaoep.mx/esucar

Tecnologías de Información
UPAEP

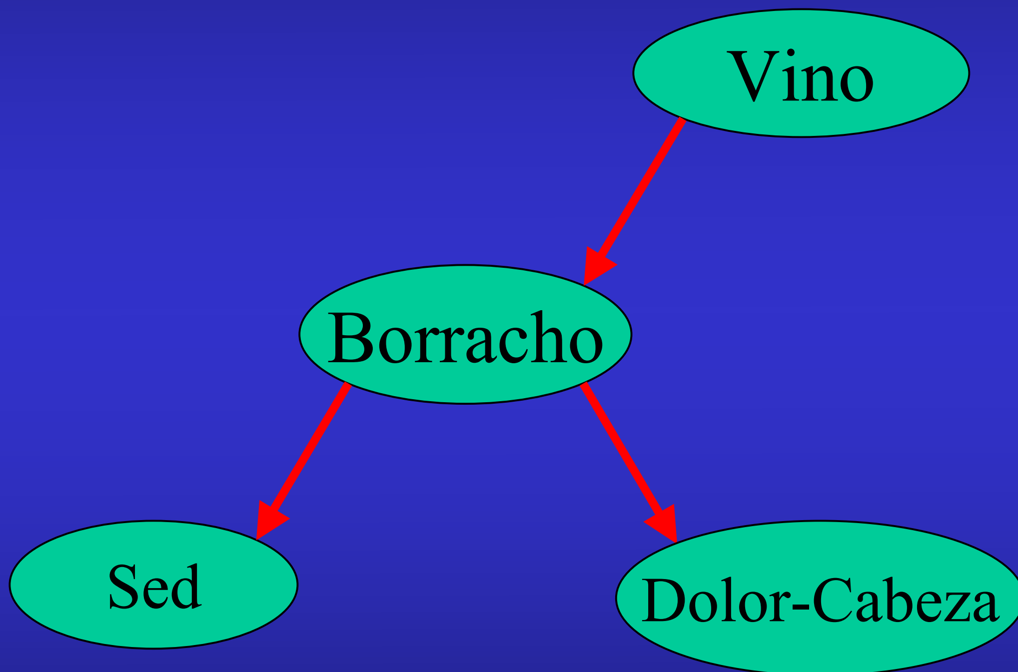
Redes Bayesianas: Parte I

- Introducción
- Representación
- Inferencia
  - Propagación en árboles

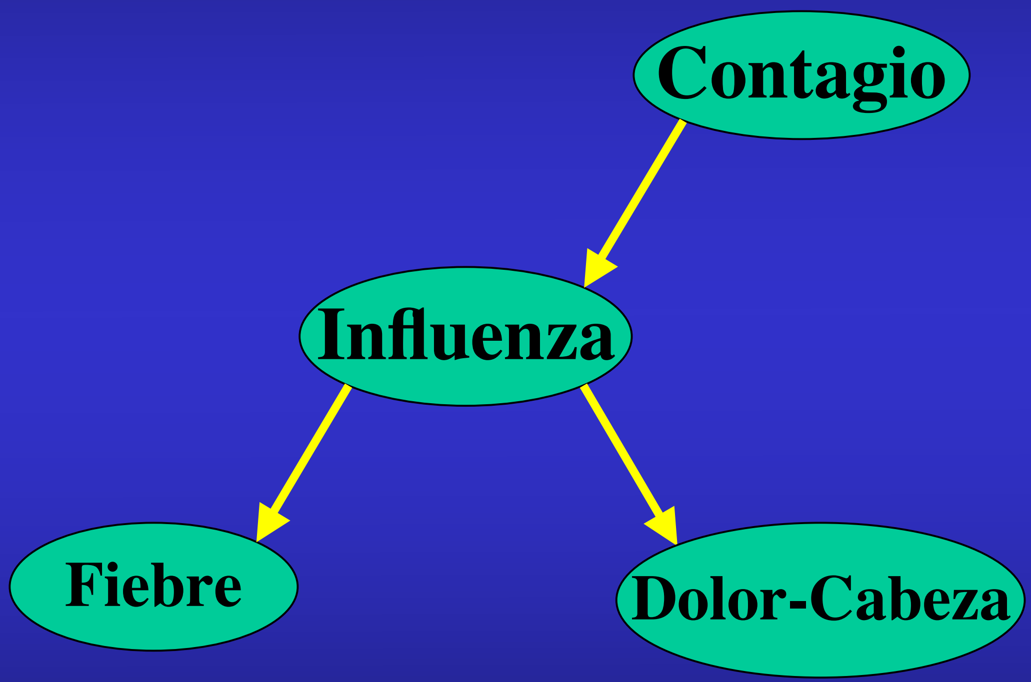
Representación

- Las redes bayesianas son una representación gráfica de dependencias para razonamiento probabilístico, en la cual los nodos y arcos representan:
  - Nodos: Variables proposicionales.
  - Arcos: Dependencia probabilística
- La variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste.

Ejemplo de una red bayesiana



Otro ejemplo …



Podemos interpretar a una RB de dos formas:

1. Distribución de probabilidad:

Representa la distribución de la probabilidad conjunta de las variables representadas en la red.

Por ejemplo:

$$P(C, I, F, D) = P(C)\, P(I \mid C)\, P(F \mid I)\, P(D \mid I)$$

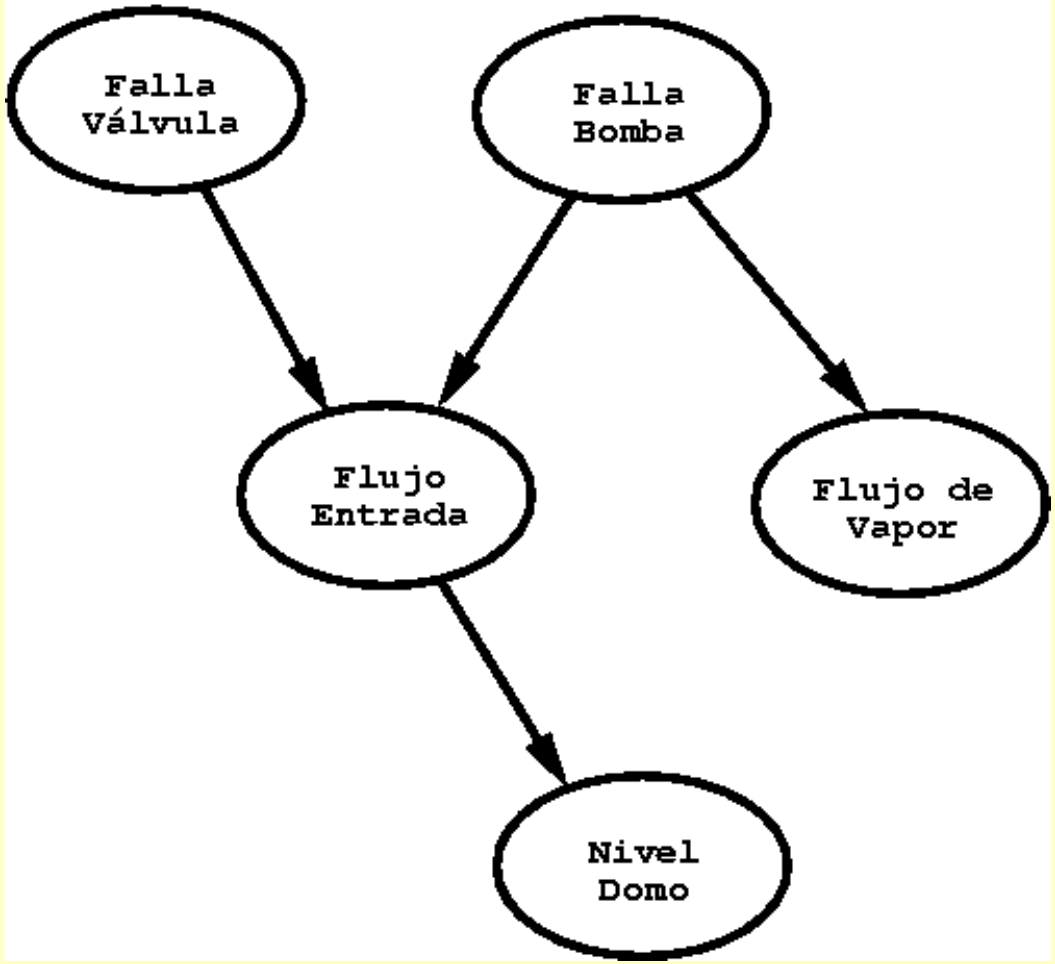
2. Base de reglas:

Cada arco representa un conjunto de reglas que asocian las variables involucradas,
Por ejemplo:

Si *I entonces F*

Dichas reglas están cuantificadas por las probabilidades respectivas.

Otro ejemplo



Estructura

- La topología o estructura de la red nos da información sobre las dependencias probabilísticas entre las variables.
- La red también representa las independencias condicionales de una variable (o conjunto de variables) dada otra variable(s).

Ejemplo

- Para el caso del domo:

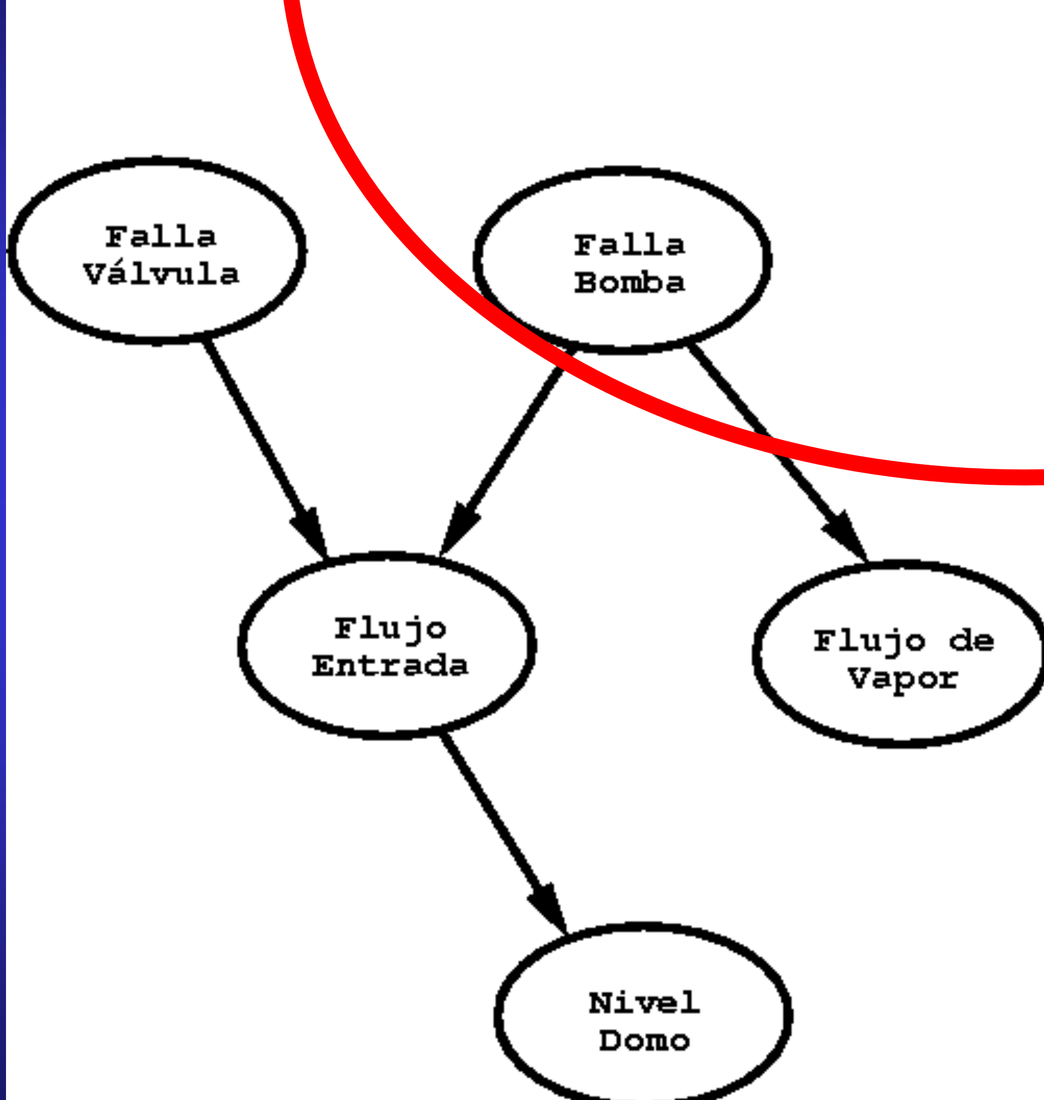
{Fva} es cond. indep. de {Fv, Fe, Nd} dado {Fb}

- Esto es:

$$P(Fva \mid Fv, Fe, Nd, Fb) = P(Fva \mid Fb)$$

- Esto se representa gráficamente por el nodo Fb separando al nodo Fva del resto de las variables.

Ejemplo de Red Bayesiana



Independencias condicionales

- En una RB todas la relaciones de independencia condicional representadas en el grafo corresponden a relaciones de independencia en la distribución de probabilidad.

- Dichas independencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades).

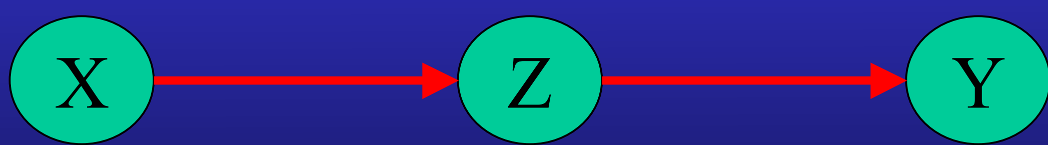
Representación Gráfica

- Una red bayesiana representa en forma gráfica las dependencias e independencias entre variables aleatorias, en particular las independencias condicionales
- Independencia en la distribución
  – $P(X \mid Y,Z) = P(X \mid Z)$
- Independencia en el grafo
  – X "separada" de Y por Z

Representación Gráfica

Notación:

- Independencia en la distribución
  - I(X,Z,Y)
- Independencia en el grafo
  - < X | Z | Y >



Separación “D”

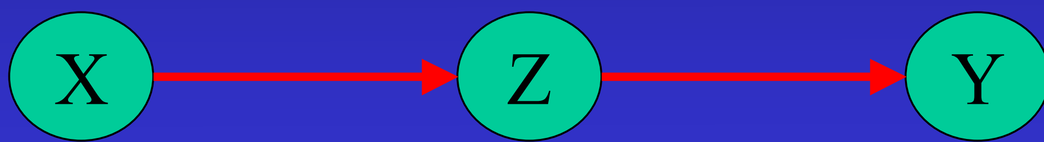
- El conjunto de variables A es independiente del conjunto B dado el conjunto C, si no existe trayectoria entre A y B en que
  1. Todos los nodos convergentes están o tienen descendientes en C
  2. Todos los demás nodos están fuera de C

Separación “D”

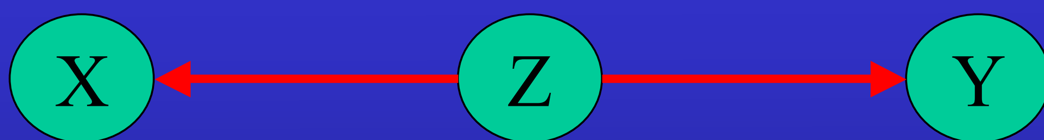
- Tres casos básicos
  - Arcos divergentes
  - Arcos en secuencia
  - Arcos convergentes

Separación "D" – casos básicos

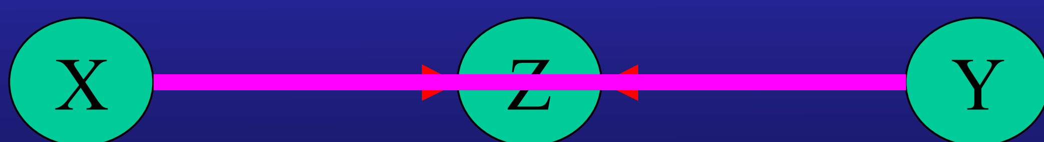
- caso 1: Secuencia:



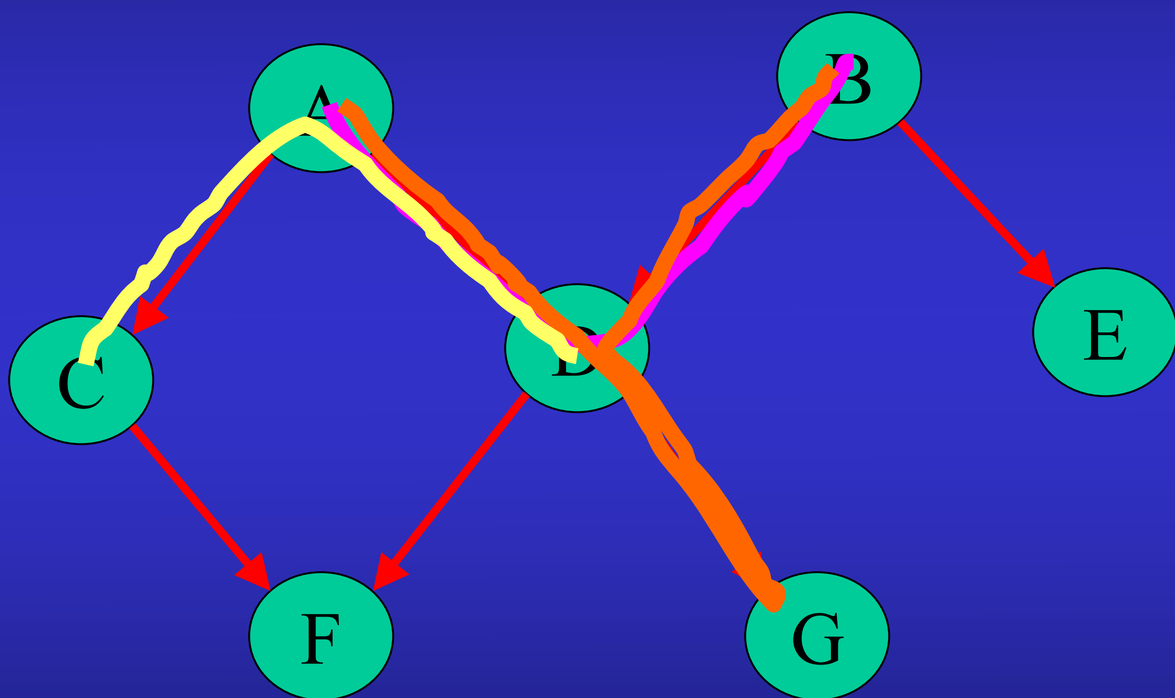
- caso 2: Divergentes:



- caso 3: Convergentes:



Ejemplos Separación-D

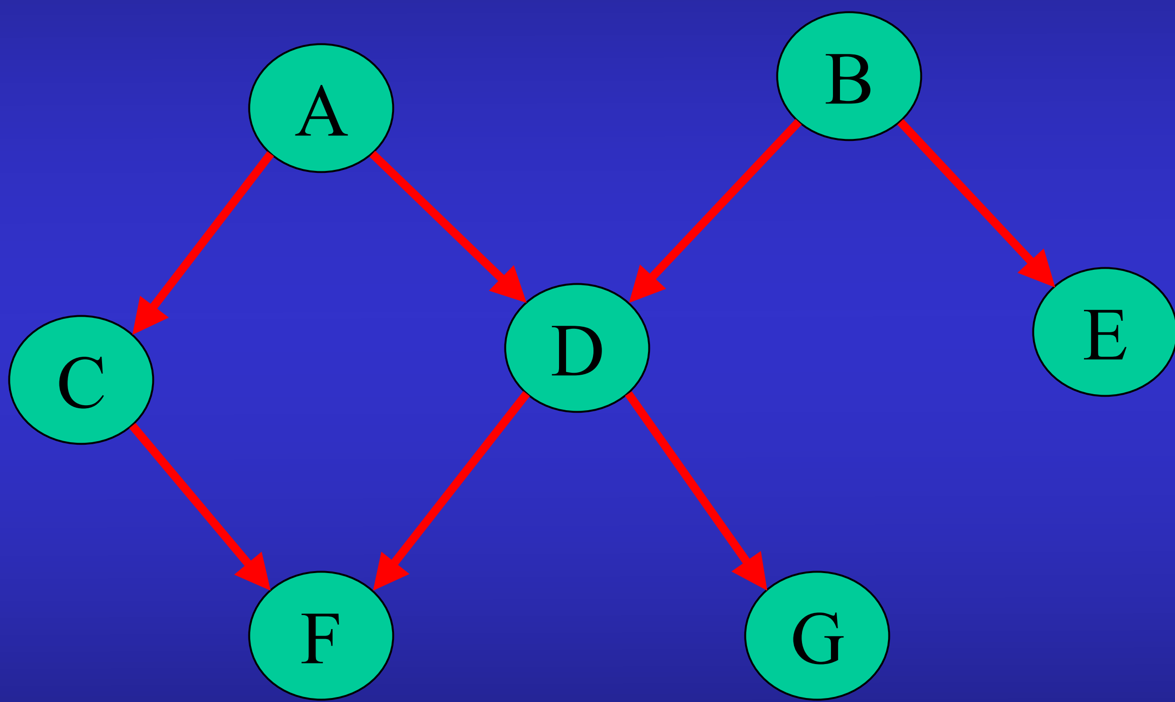


¿I(A,CD,F)?
¿I(A,CD,B)?
¿I(BD,A,C)?
¿I(A,G,B)?
¿I(A,D,G)?
¿I(C,BEG,D)?

Especificación Estructural

- En una RB, cualquier nodo X es independiente de todos los nodos que no son sus descendientes dados sus nodos padres Pa(X) – “contorno de X”
- La estructura de una RB se especifica indicando el contorno (padres) de cada variable

Especificación Estructural

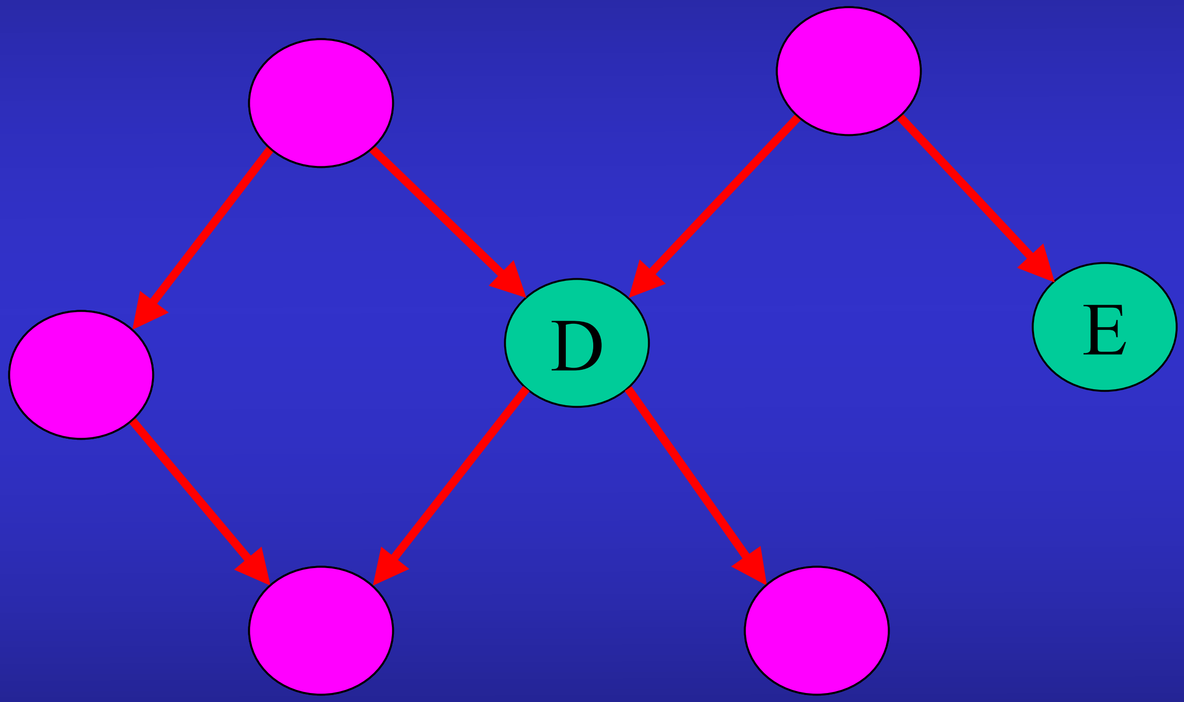


Pa(A) = 0
Pa(B) = 0
Pa(C) = A
Pa(D) = A, B
Pa(E) = B
Pa(F) = C, D
Pa(G) = D

Cobija de Markov

- La “cobija de Markov” de un nodo es el conjunto de nodos que lo hacen independiente del resto de la red
- Para una RB la cobija de Markov está formada por:
  - Nodos padre
  - Nodos hijo
  - Otros padres de los hijos

Cobija de Markov

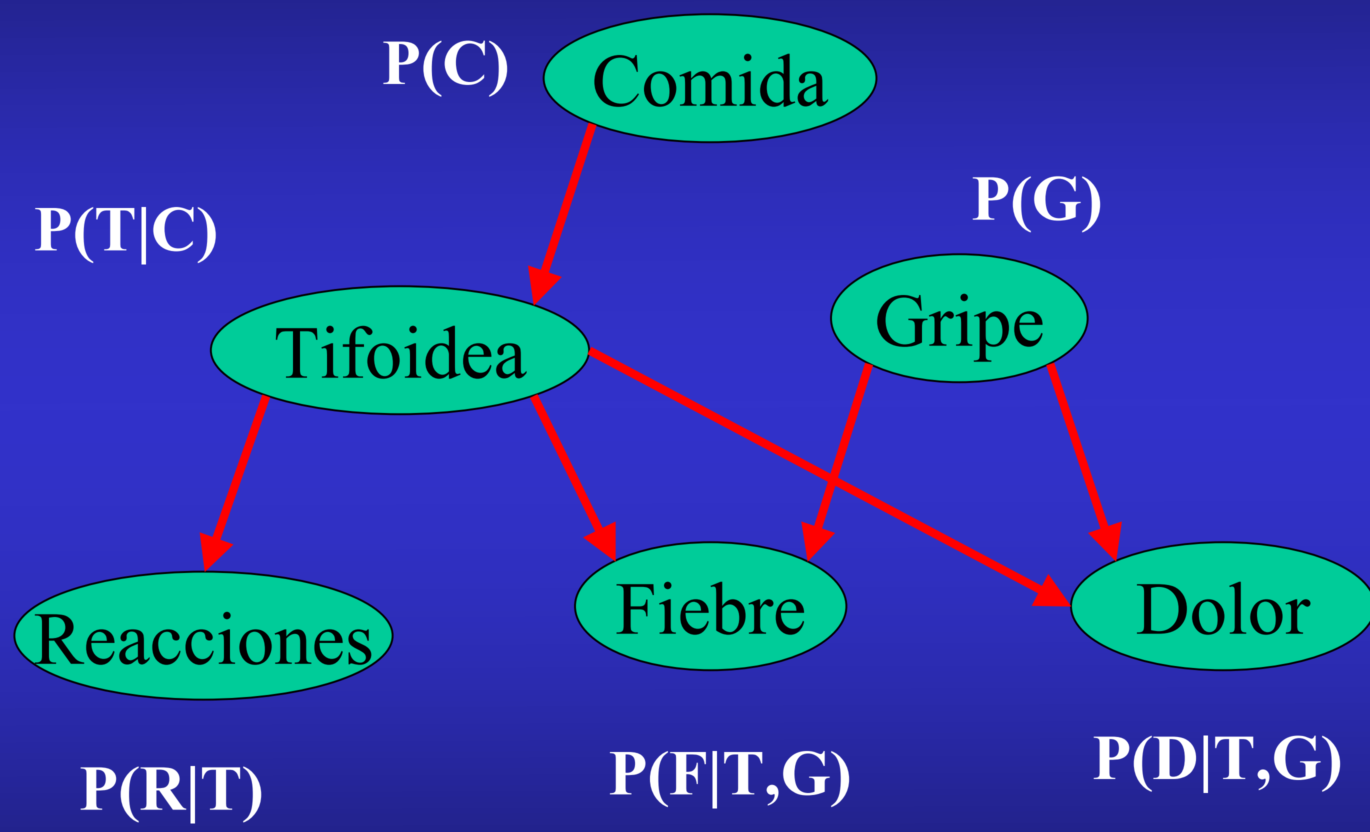


CM (D) ?

Parámetros

- Complementan la definición de una red bayesiana las probabilidades condicionales de cada variable dados sus padres.
  - **Nodos raíz**: vector de probabilidades marginales
  - **Otros nodos**: matriz de probabilidades condicionales dados sus padres

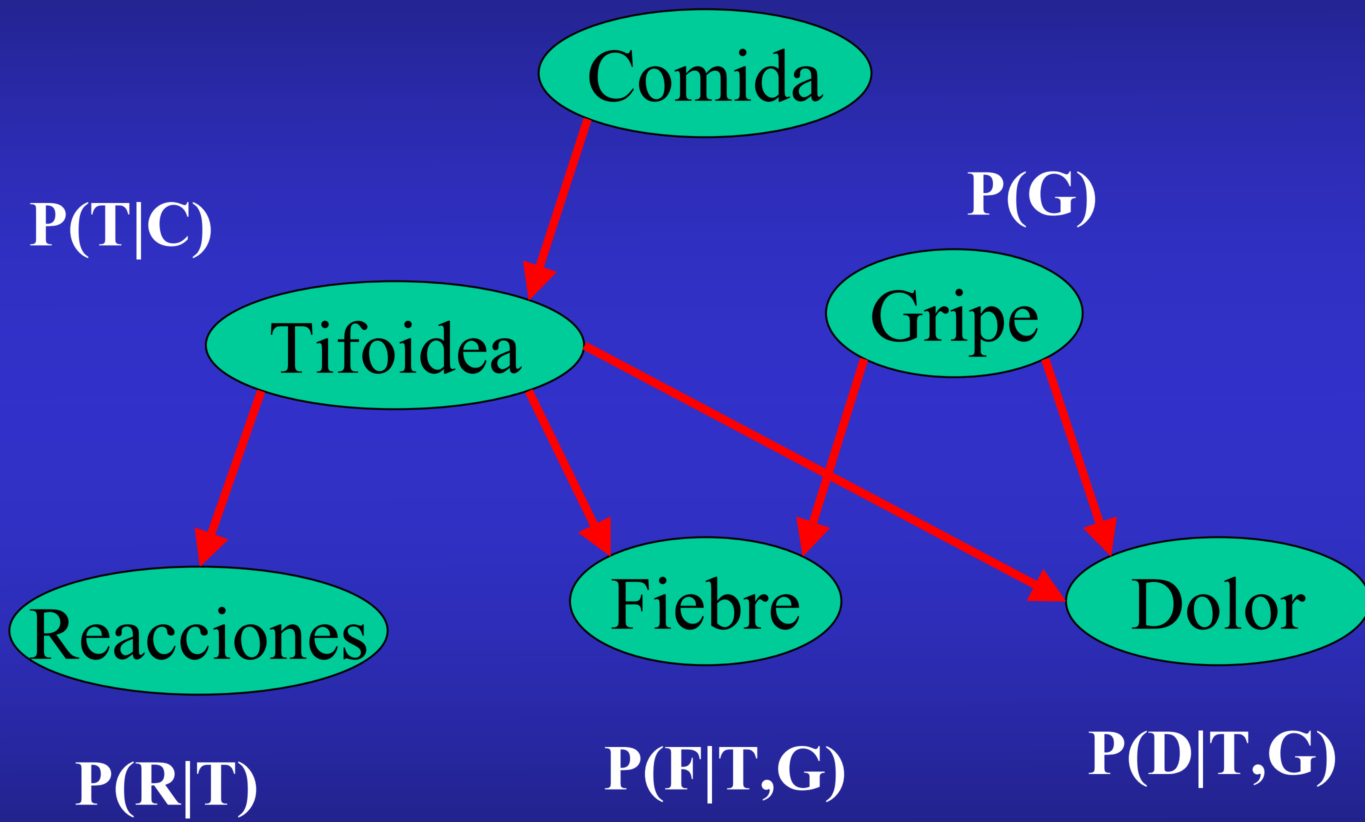
Ejemplo


P(C)
Comida
P(T|C)
P(G)
Tifoidea
Gripe
Reacciones
Fiebre
Dolor
P(R|T)
P(F|T,G)
P(D|T,G)

Ejemplo

P(C)

Ins	Sal
0.2	0.8

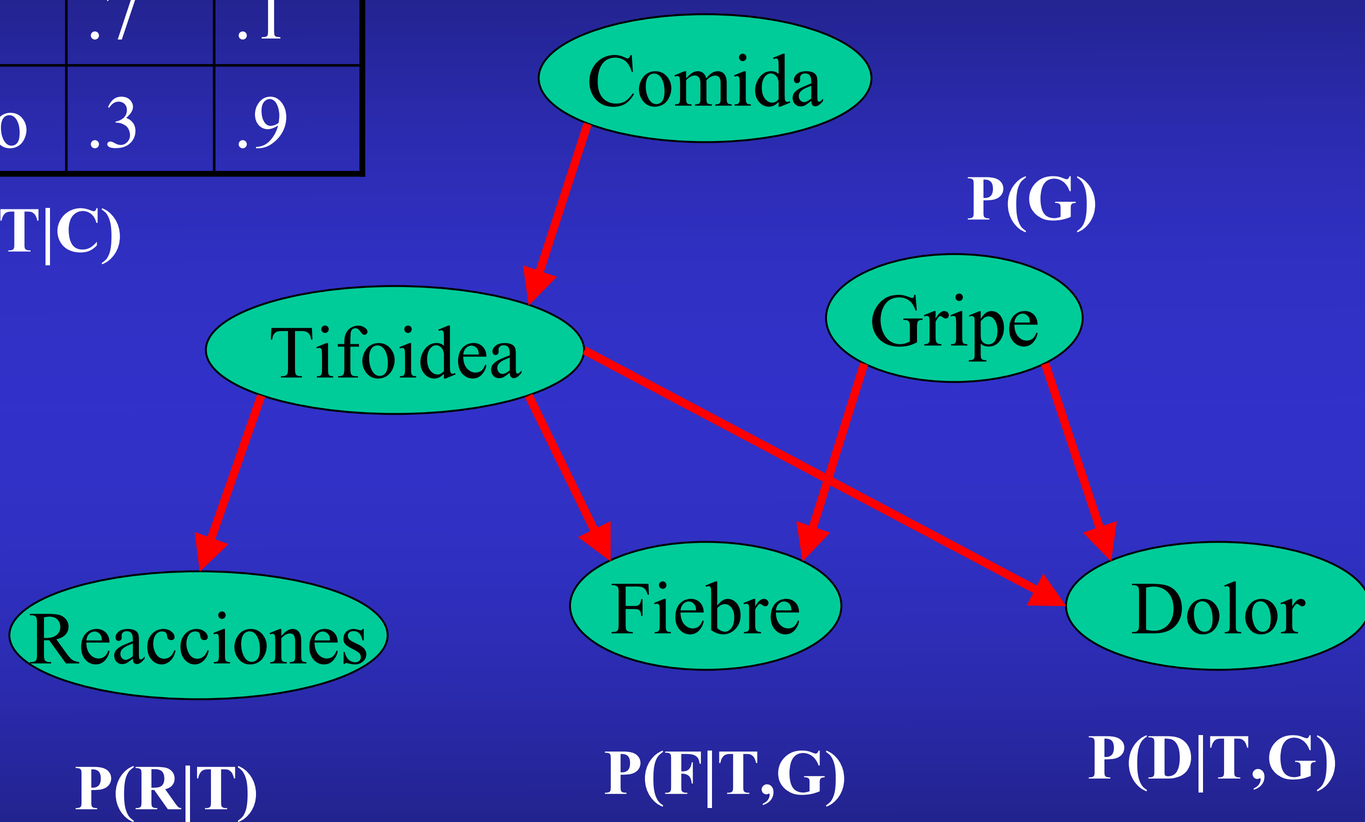


P(C)

Ins	Sal
0.2	0.8

	Ins	Sal
Si	.7	.1
No	.3	.9

P(T|C)

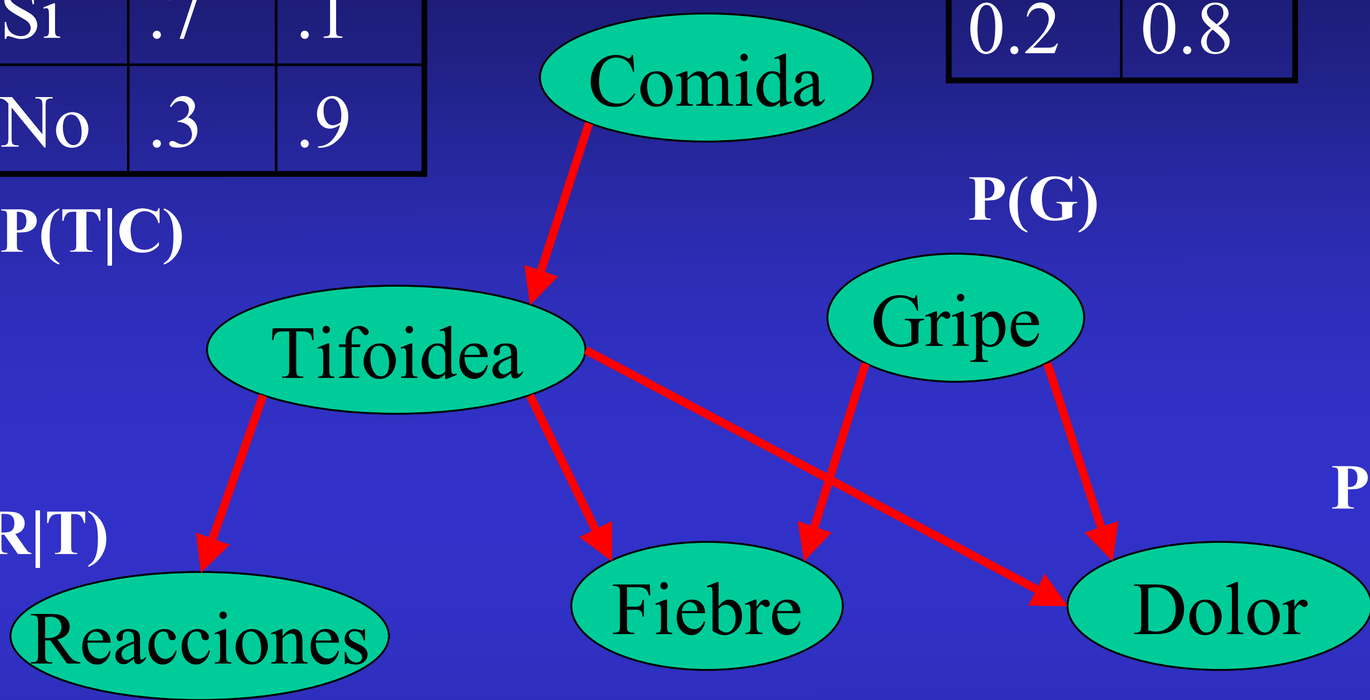


	Ins	Sal
Si	.7	.1
No	.3	.9

P(T|C)

P(C)

Ins	Sal
0.2	0.8



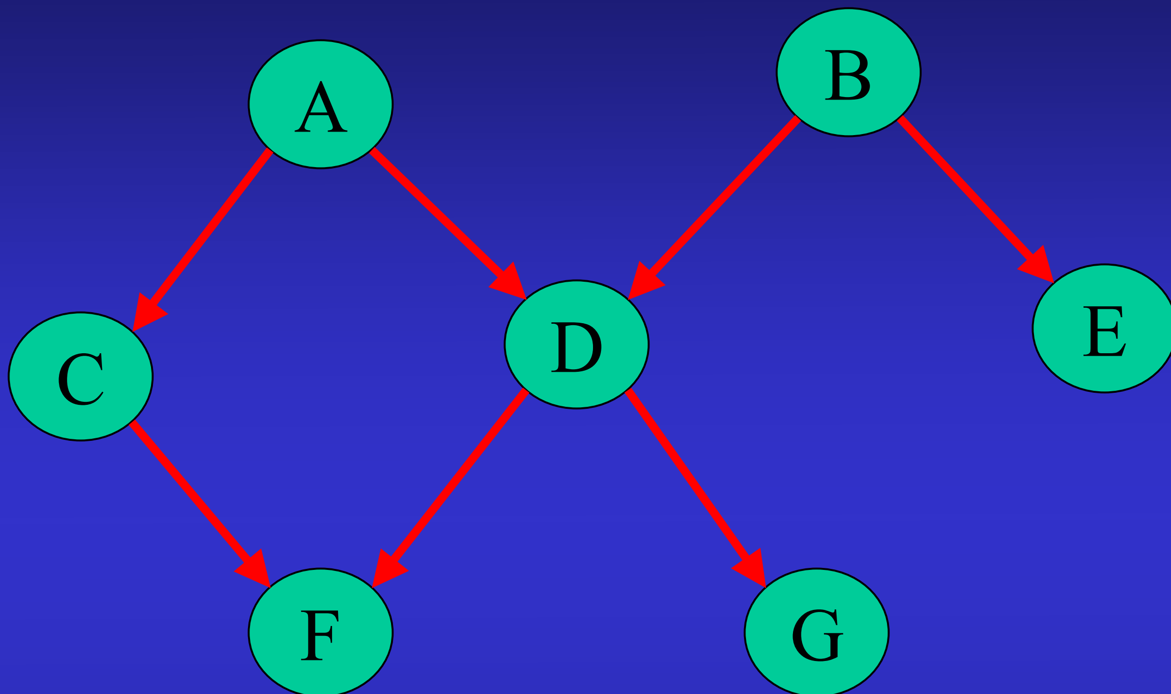
P(F|T,G)

	Si, Si	Si,No	No,Si	No,No
F	0.8	0.6	0.5	0.1
~F	0.2	0.4	0.5	0.9

Especificación Paramétrica

- Dado que los contornos (padres) de cada nodo especifican la estructura, mediante las probabilidades condicionales de dichos nodos podemos especificar también las probabilidades requeridas
- Aplicando la regla de la cadena y las independencias condicionales, se puede verificar que con dichas probabilidades se puede calcular la probabilidad conjunta

Especificación Paramétrica



$P(A,B,C,D,E,F,G)$

$= \quad P(G|F,E,D,C,B,A)\ P(F|E,D,C,B,A)\ P(E|D,C,B,A)\ P(D|C,B,A)\ P(C|B,A)\ P(B|A)\ P(A)$

$= \quad P(G|D)\ P(F|D,C)\ P(E|B)\ P(D|B,A)\ P(C|A)\ P(B)\ P(A)$

Especificación Paramétrica

- En general, la probabilidad conjunta se especifica por el producto de las probabilidades de cada variable dados sus padres:

$$P(X1,X2, ..., Xn) = \prod P(Xi \mid Pa(Xi))$$

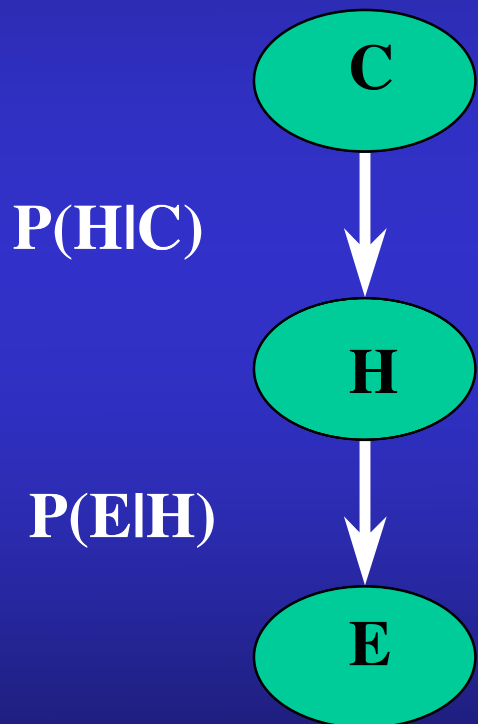
Inferencia probabilística

- En RB, la inferencia probabilística consiste en:

"dadas ciertas variables conocidas (evidencia), calcular la probabilidad posterior de las demás variables (desconocidas)"

- Es decir, calcular: P(Xi | **E**), donde:
  - **E** es un subconjunto de variables de la RB (posiblemente vació)
  - Xi es cualquier variable en la RB, no en **E**

Inferencia bayesiana



Causal:

C → H

Evidencial:

E → H

Mixta:

C, E → H

Tipos de Técnicas

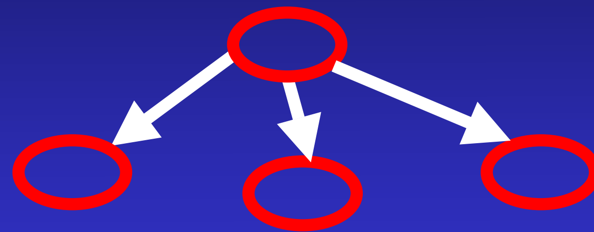
- Calcular probabilidades posteriores:
  - Una variable, cualquier estructura: algoritmo de eliminación (*variable elimination*)
  - Todas las variable, estructuras sencillamente conectadas (árboles, poliárboles): propagación
  - Todas las variables, cualquier estructura:
    - Agrupamiento (*junction tree*)
    - Simulación estocástica
    - Condicionamiento

Tipos de Técnicas

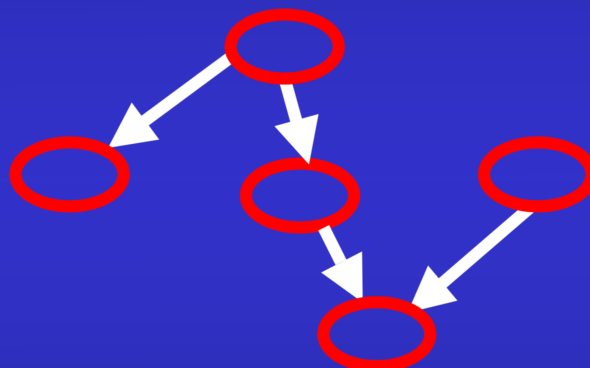
- Obtener variable(s) de mayor probabilidad dada cierta evidencia – abducción:
  - Abducción total
  - Abducción parcial

Tipos de estructuras

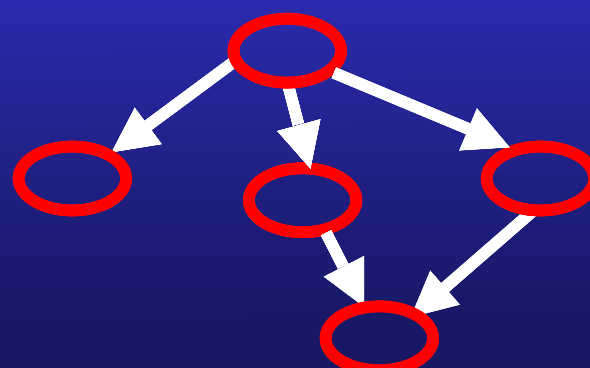
- Sencillamente conectadas
  - Árboles



  - Poliárboles



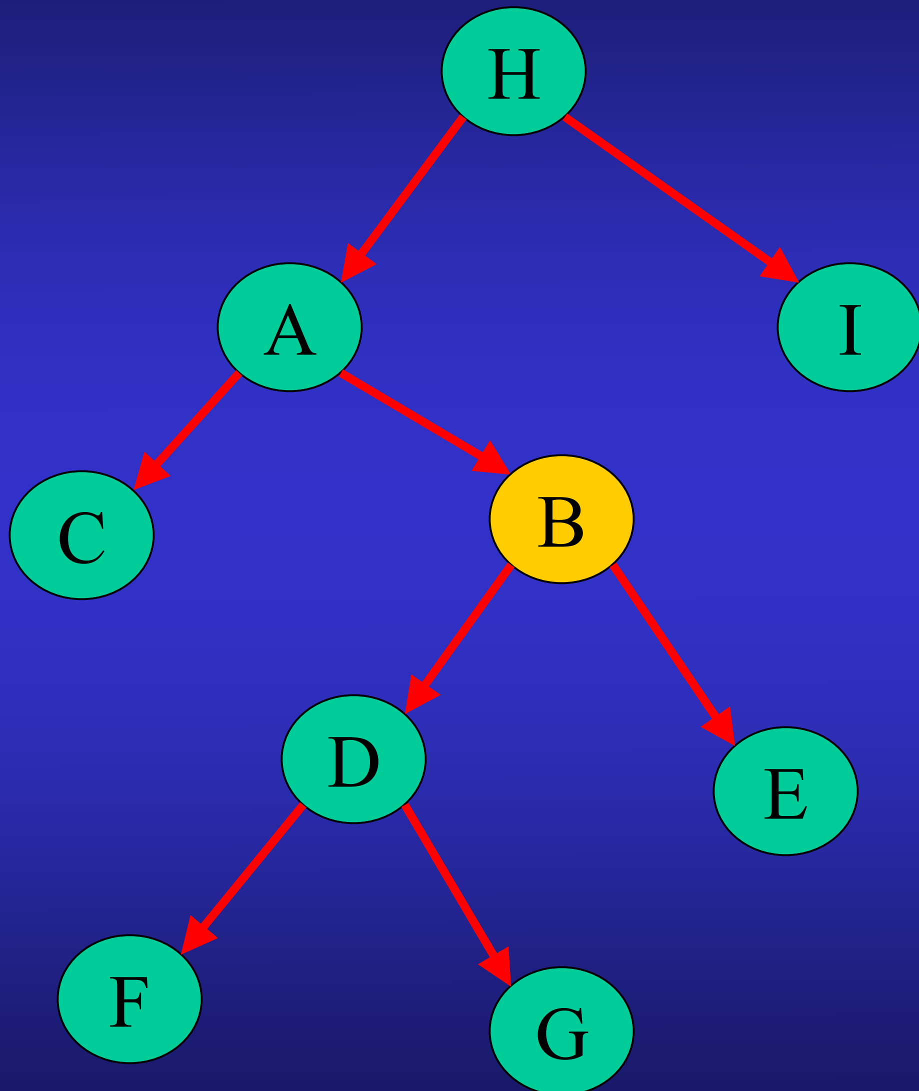
- Multiconectadas



Propagación en Árboles

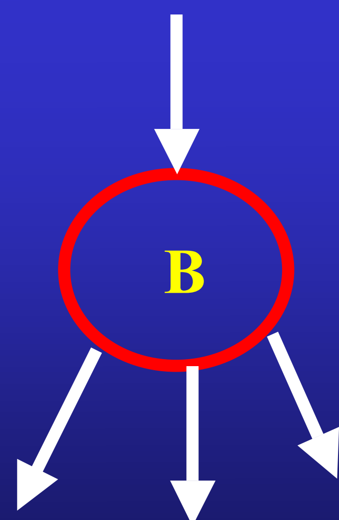
Cada nodo corresponde a una variable discreta, B (B_1, B_2,…, B_m) con su respectiva matriz de probabilidad condicional, $P(B|A)=P(B_j|\ A_i)$

Propagación en Árboles

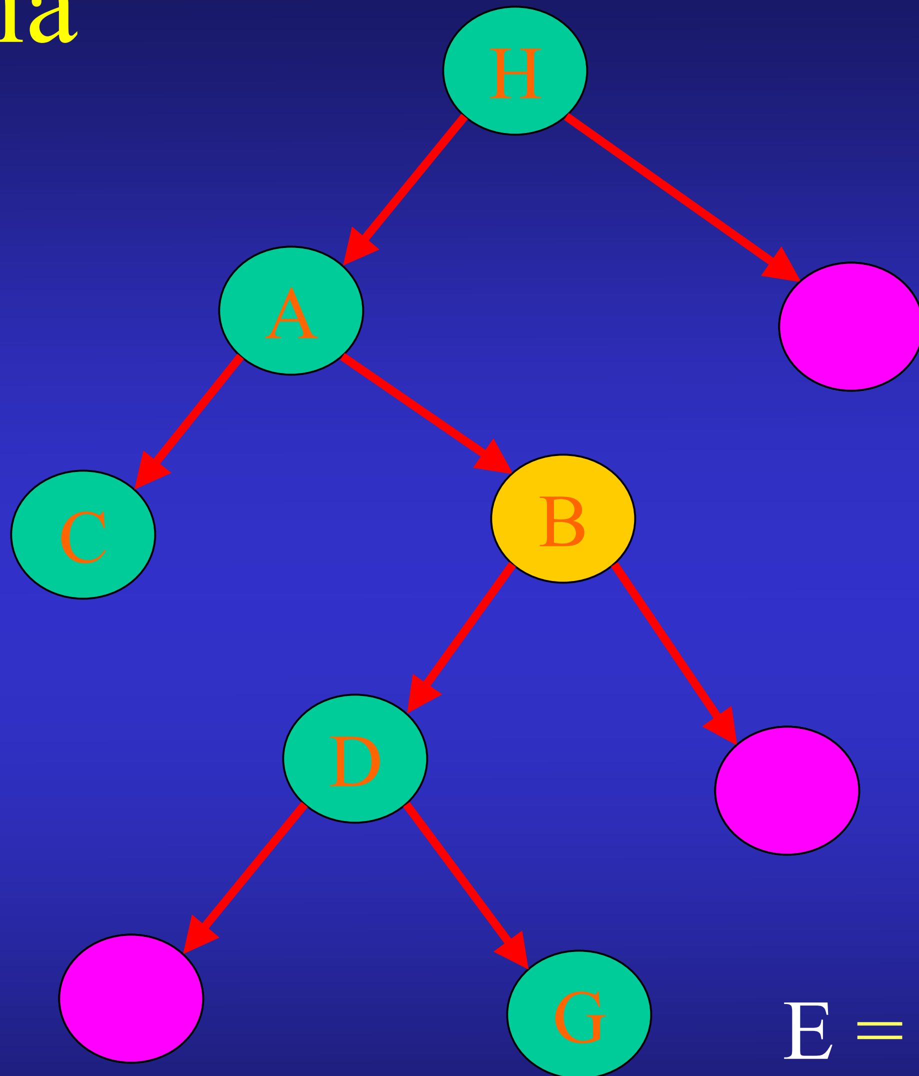


Dada cierta evidencia ***E*** -representada por la instanciación de ciertas variables- la probabilidad posterior de cualquier variable ***B***, por el teorema de Bayes:

$$P(B_i \mid E) = P(B_i)\, P(E \mid B_i) / P(E)$$



Evidencia



E = {I,F,E}

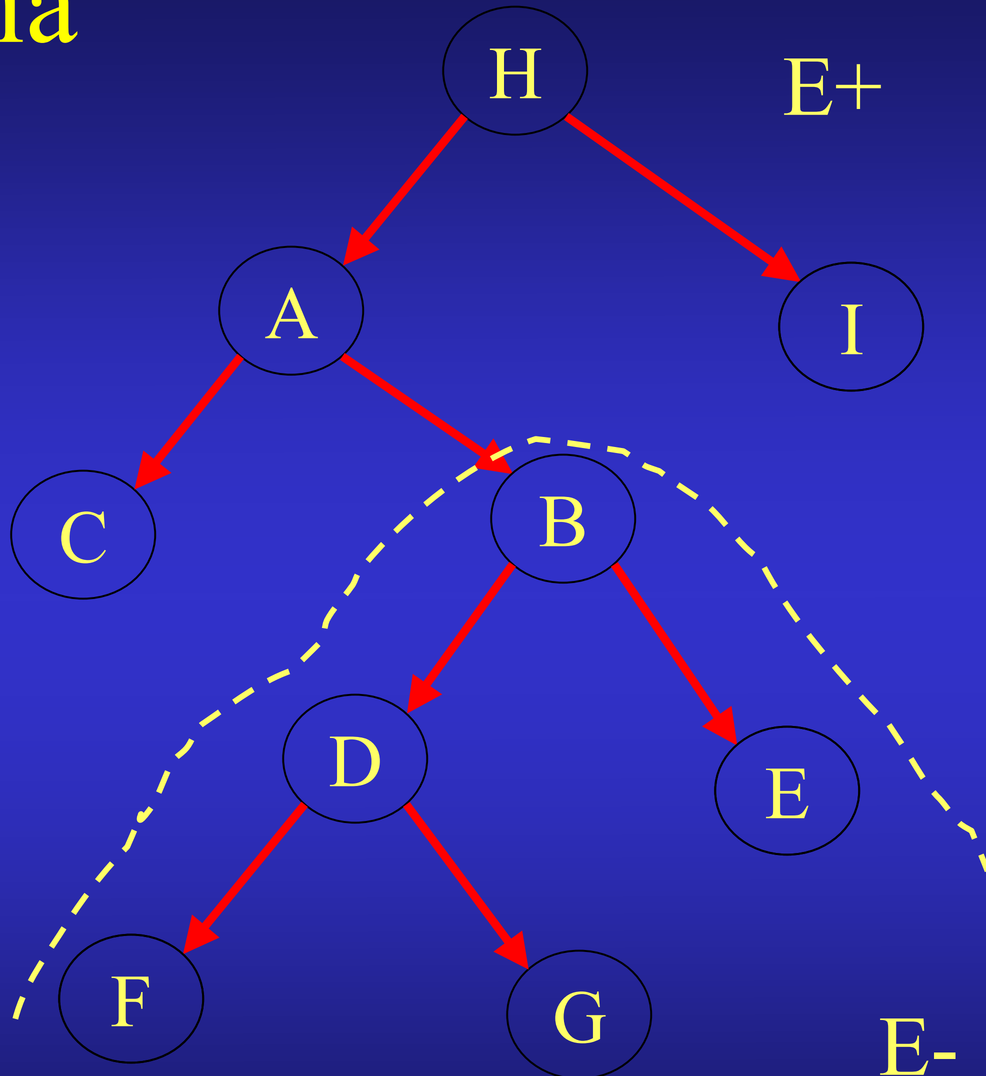
Evidencia

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:

E-: Datos en el árbol que cuya raíz es ***B***

E+: Datos en el resto del árbol

Evidencia



Entonces:

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

Pero dado que ambos son independientes y aplicando nuevamente Bayes:

$$P(B_i \mid E) = \alpha\, P(B_i \mid E^+)\, P(E^- \mid B_i)$$

Donde α es una constante de normalización

Definiciones:

Si definimos los siguientes términos:

$$\lambda\,(B_i) = P\,(\,E^- \mid B_i)$$

$$\pi\,(B_i) = P\,(B_i \mid E^+\,)$$

Entonces:

$$P(B_i \mid E\,) = \alpha\,\pi\,(B_i)\,\lambda\,(B_i)$$

Desarrollo

- En base a la ecuación anterior, se puede integrar un algoritmo distribuido para obtener la probabilidad de un nodo dada cierta evidencia
- Para ello se descompone el cálculo de cada parte:
  - Evidencia de los hijos (λ)
  - Evidencia de los demás nodos (π)

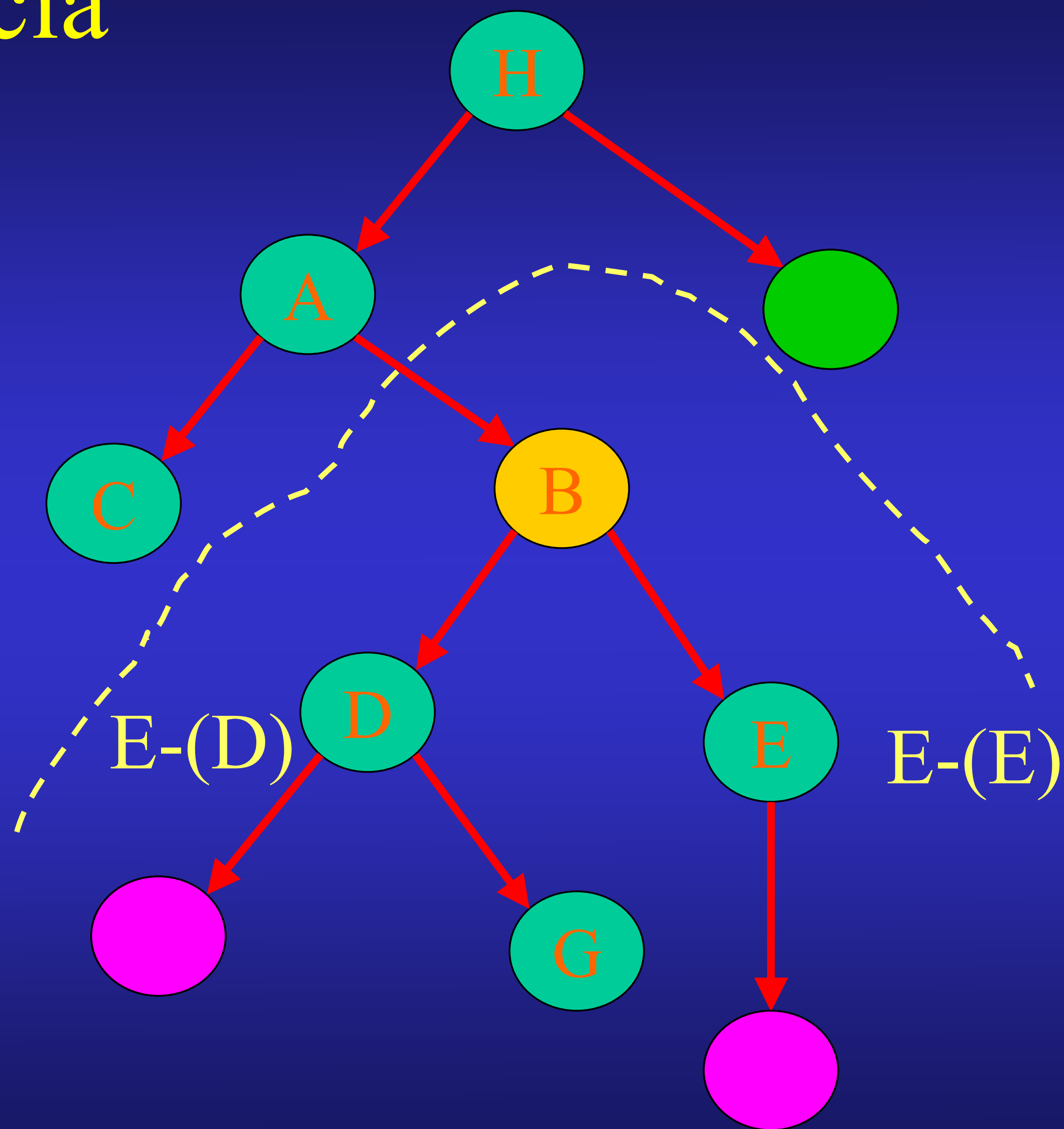
Evidencia de los hijos (λ)

- Dado que los hijos son condicionalmente independientes dado el padre:

$$\lambda\ (B_i) = P\ (\ E^-\ |\ B_i) = \Pi_k\ P\ (\ E_k^-\ |\ B_i)$$

- Donde E_k^- corresponde a la evidencia del subárbol del hijo k

Evidencia hijos



Evidencia de los hijos (λ)

- Condicionando respecto a los posibles valores de los hijos de B:

$$\lambda\,(B_i) = \Pi_k \left[\, \Sigma_j\, P\,(\,E_k^- \mid B_i, S_j^k)\, P(S_j^k \mid B_i)\,\right]$$

- Donde S^k es el hijo k de B, y la sumatoria es sobre los valores de dicho nodo (teorema de probabilidad total)

Evidencia de los hijos (λ)

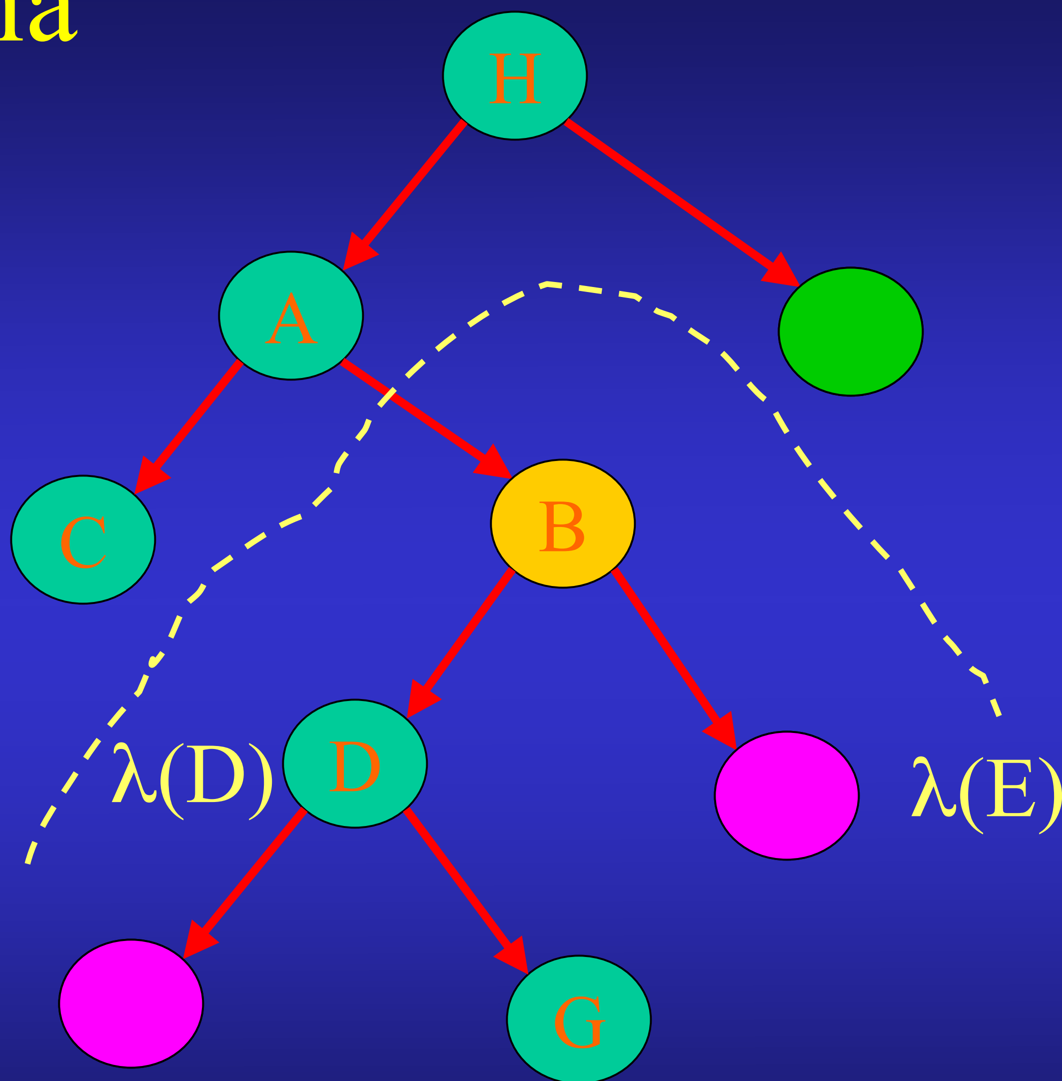
- Dado que B es condicionalmente independiente de la evidencia dados sus hijos:

$$\lambda\,(B_i) = \Pi_k\,[\,\Sigma_j\,P\,(\,E_k^-\,|\,S_j^k)\,P(S_j^k\,|\,B_i)\,]$$

- Substituyendo la definción de λ:

$$\lambda\,(B_i)= \Pi_k\,[\,\Sigma_j\,P(S_j^k\,|\,B_i)\,\lambda\,(S_j^k)]$$

Evidencia hijos



Evidencia de los hijos (l)

- Recordando que λ es un vector (un valor por cada posible valor de B), lo podemos ver en forma matricial:

$$\boxed{\lambda} = \boxed{\lambda}\ \boxed{P\,(S \mid B)}$$

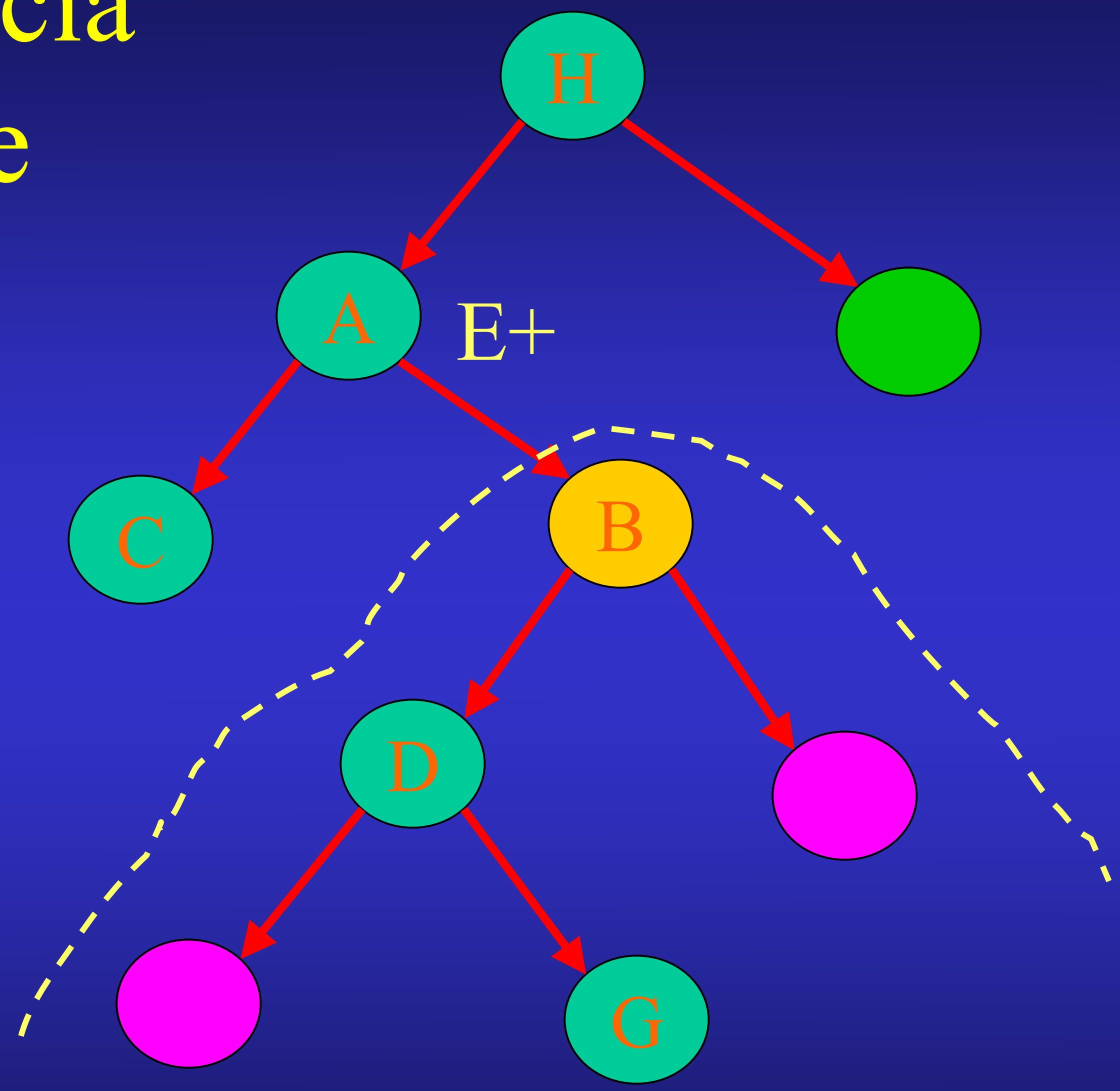
Evidencia de los demás nodos (π)

- Condicionando sobre los diferentes valores del nodo padre (A):

$$\pi\ (B_i) = P\ (B_i \mid E^+) = \Sigma_j\ P\ (B_i \mid E^+, A_j)\ P(A_j \mid E^+)$$

- Donde A_j corresponde a los diferentes valores del nodo padre de B

Evidencia padre



Evidencia de los demás nodos (p)

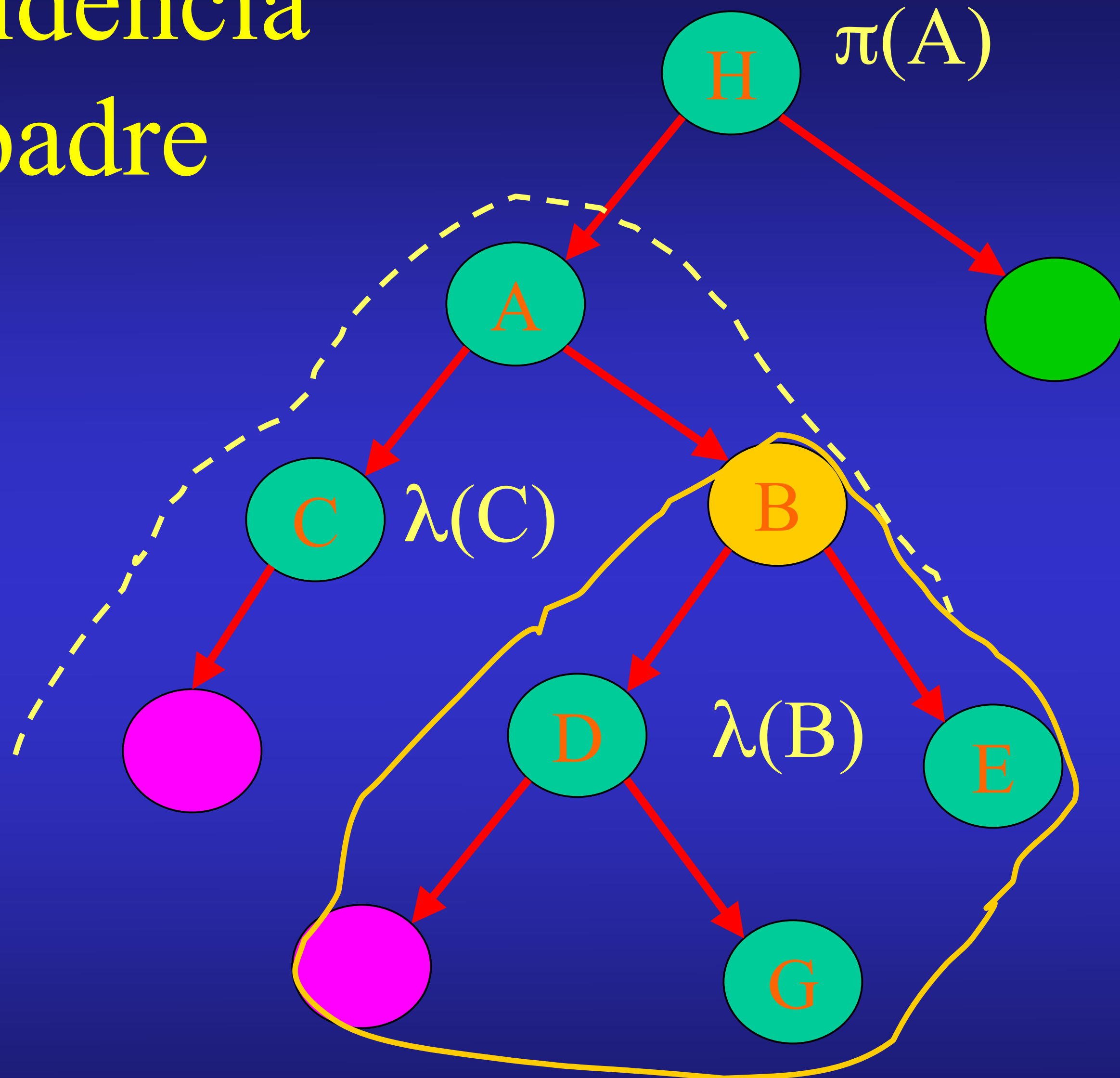
- Dado que B es independiente de la evidencia "arriba" de A, dado A:

$$\Pi\ (B_i) = \Sigma_j\ P\ (B_i \mid A_j)\ P(A_j \mid E^+)$$

- La $P(A_j \mid E^+)$ corresponde a la P posterior de A dada toda la evidencia excepto B y sus hijos, por lo que se puede escribir como:

$$P(A_j \mid E^+) = \alpha\ \pi\ (A_i)\ \Pi_{k^1 B}\ \lambda_k\ (A_i)$$

Evidencia padre



Evidencia de los demás nodos (π)

- Substituyendo $\mathbf{P(A_j \mid E^+)}$ en la ecuación de π :

$$\pi\ (B_i) = \sum_j P\ (B_i \mid A_j) \left[\ \alpha\ \pi\ (A_i)\ \prod_{k \neq B} \lambda_k\ (A_i)\ \right]$$

- De forma que se obtiene combinando la π de del nodo padre con la λ de los demás hijos

Evidencia de los demás nodos (p)

- Dado que también π es un vector, lo podemos ver en forma matricial (donde P_A es el producto de la evidencia de padre y otros hijos):

$$\boxed{\pi} = \boxed{P(B \mid A)} \; \boxed{P_A}$$

Algoritmo

- Mediante estas ecuaciones se integra un algoritmo de propagación de probabilidades en árboles.
- Cada nodo guarda los valores de los vectores π y λ, así como su matriz de probabilidad condicional (CPT), P.
- La propagación se hace por un mecanismo de paso de mensajes, en donde cada nodo envía los mensajes correspondientes a su padre e hijos

Mensaje al padre (hacia arriba) –
nodo B a su padre A:

$$\lambda_B(A_i) = \sum_j P(B_j|A_i)\lambda(B_j)$$

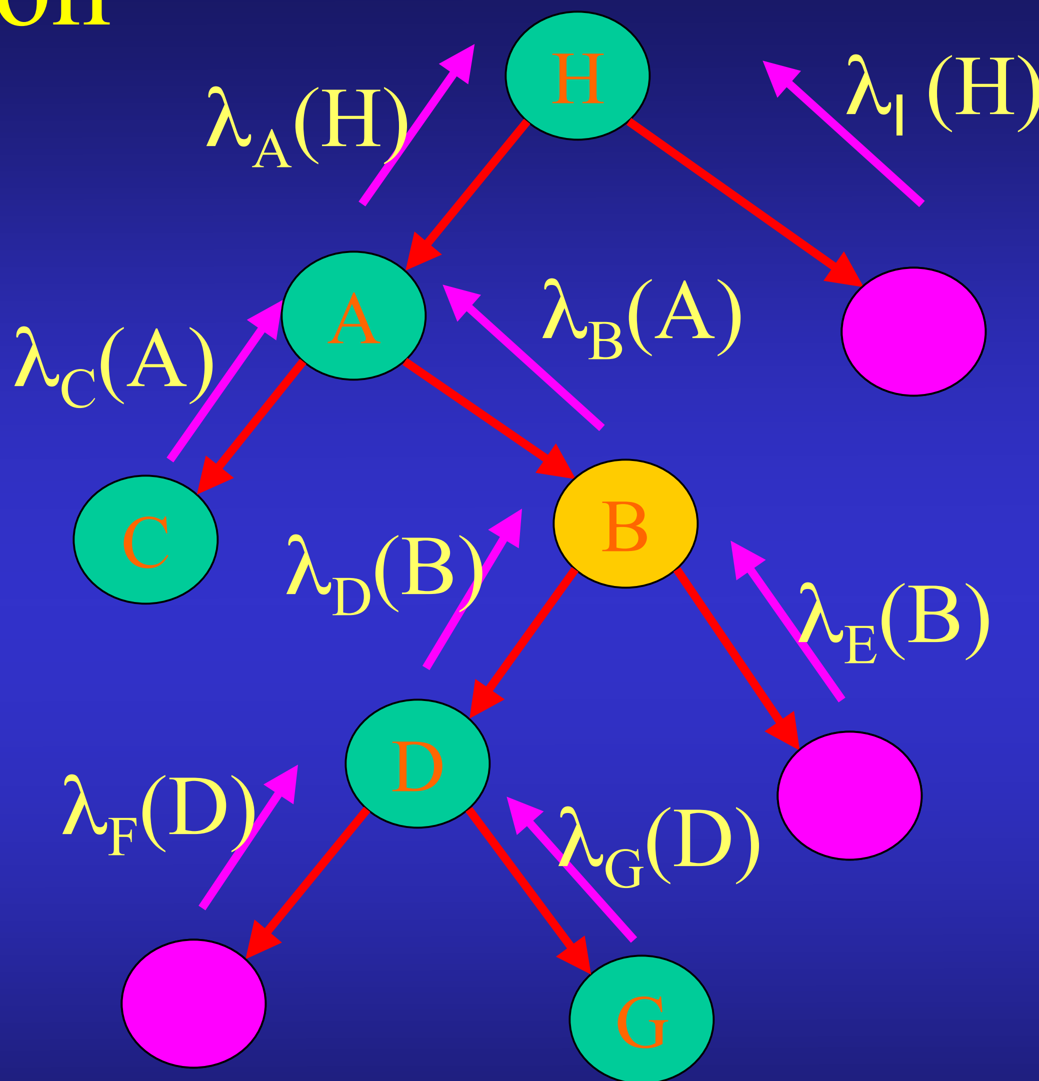
Mensaje a los hijos (hacia abajo) -
nodo B a su hijo S_k :

$$\pi_k(B_i) = \alpha\pi(B_j)\prod_{l \neq k}\lambda_l(B_j)$$

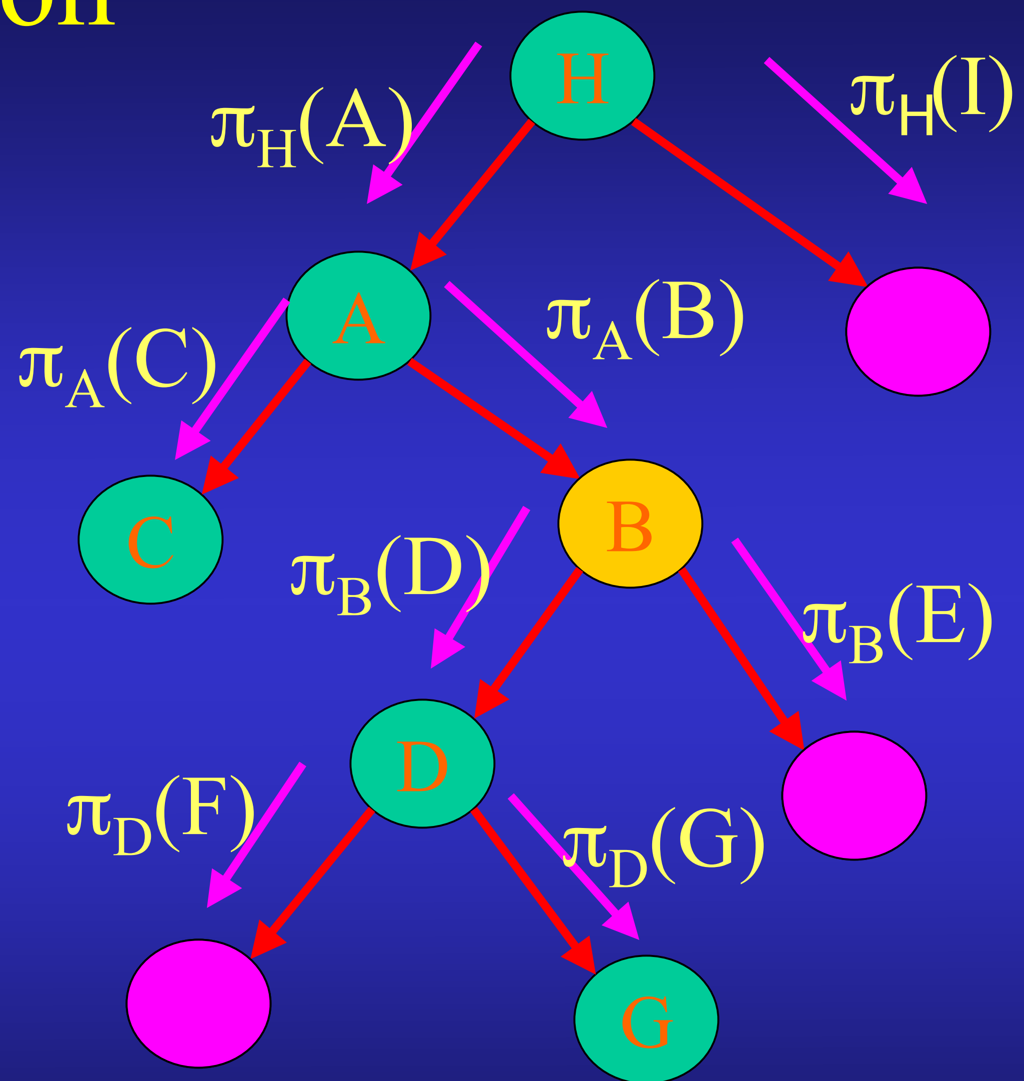
Algoritmo

- 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:

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

- Nodos asignados (conocidos):

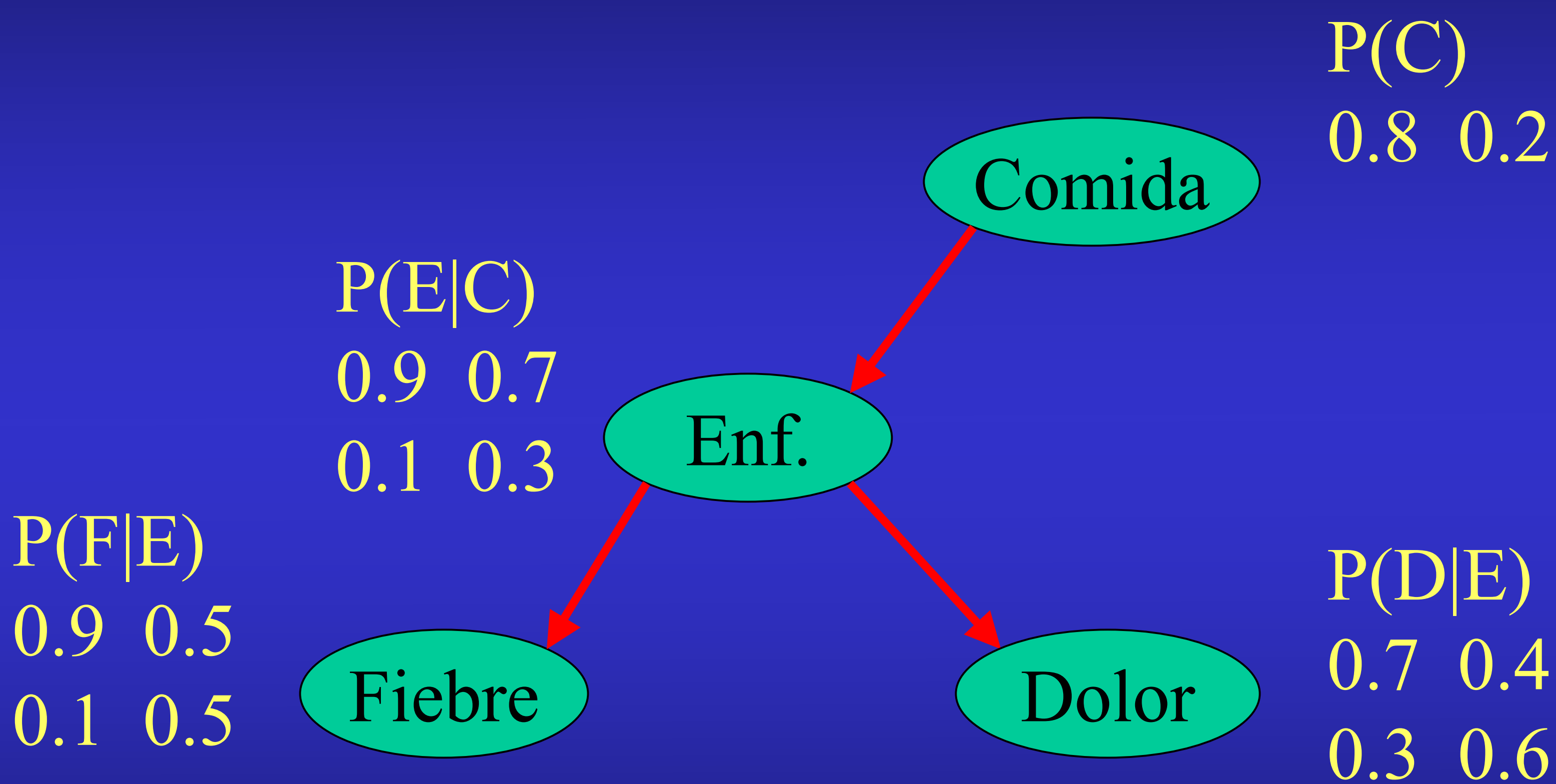
$\lambda\ (B_i) = [0,0, ..1, 0, \ldots, 0]$ (1 para valor asignado)

$\pi\ (B_i) = [0,0, ..1, 0, \ldots, 0]$ (1 para valor asignado)

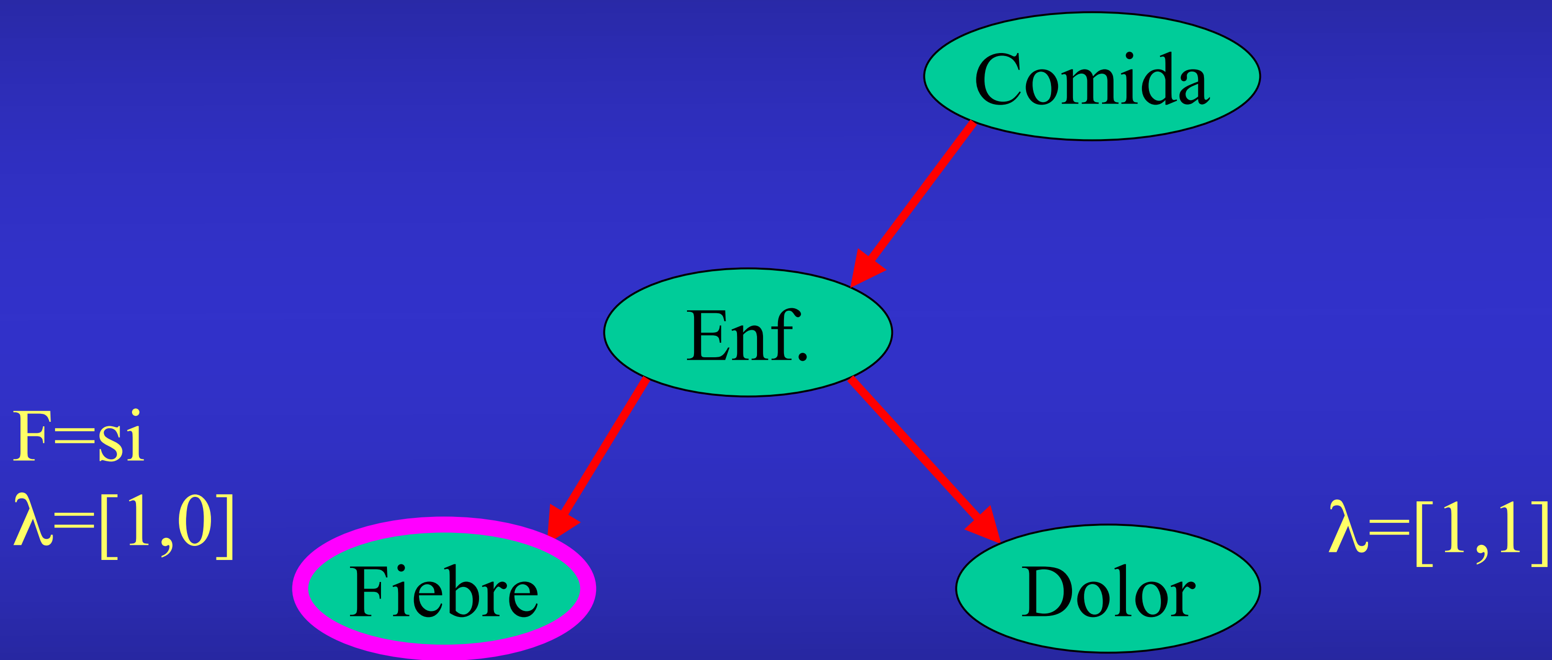
- Nodo raíz no conocido:

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

Ejemplo

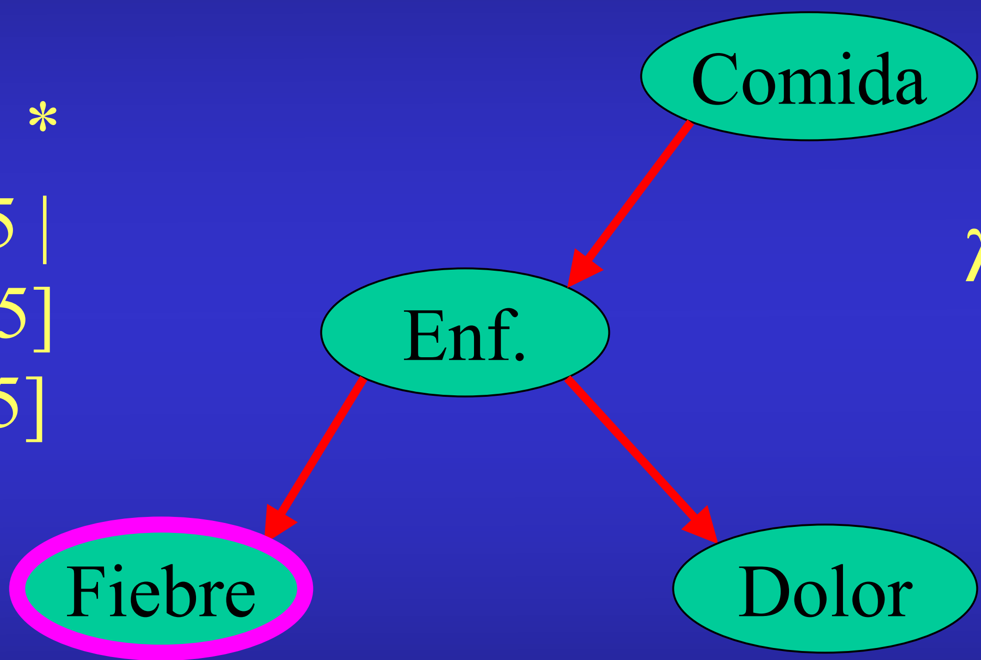

P(C)
0.8 0.2
Comida
P(E|C)
0.9 0.7
0.1 0.3
Enf.
P(F|E)
0.9 0.5
0.1 0.5
Fiebre
Dolor
P(D|E)
0.7 0.4
0.3 0.6

Ejemplo



Ejemplo

$\lambda_F = [1,0]$ *
$[.9\ .5\ |$
$.1\ .5]$
$= [.9\ .5]$



$\lambda_D = [1,1]$ *
$[.7\ .4\ |$
$.3\ .6]$
$= [1\ \ 1]$

P(F|E)
0.9 0.5
0.1 0.5

P(D|E)
0.7 0.4
0.3 0.6

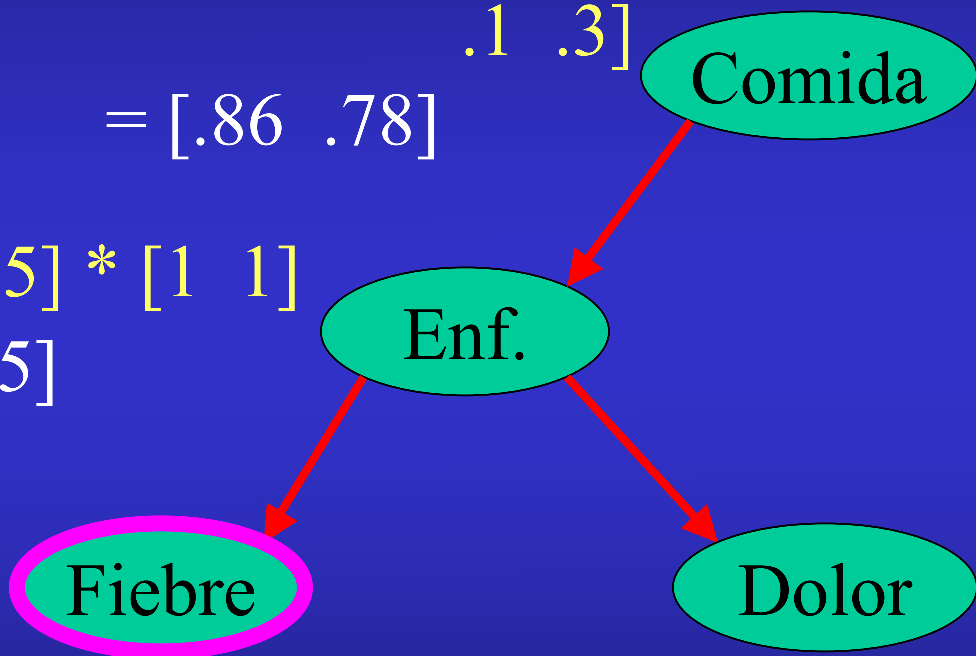
Ejemplo

$$\lambda(C) = [.9 \;\; .5] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix}$$

$$= [.86 \;\; .78]$$

$$\lambda(E) = [.9 \;\; .5] * [1 \;\; 1]$$

$$= [.9 \;\; .5]$$



P(E|C)

0.9 0.7

0.1 0.3

P(F|E)

0.9 0.5

0.1 0.5

P(D|E)

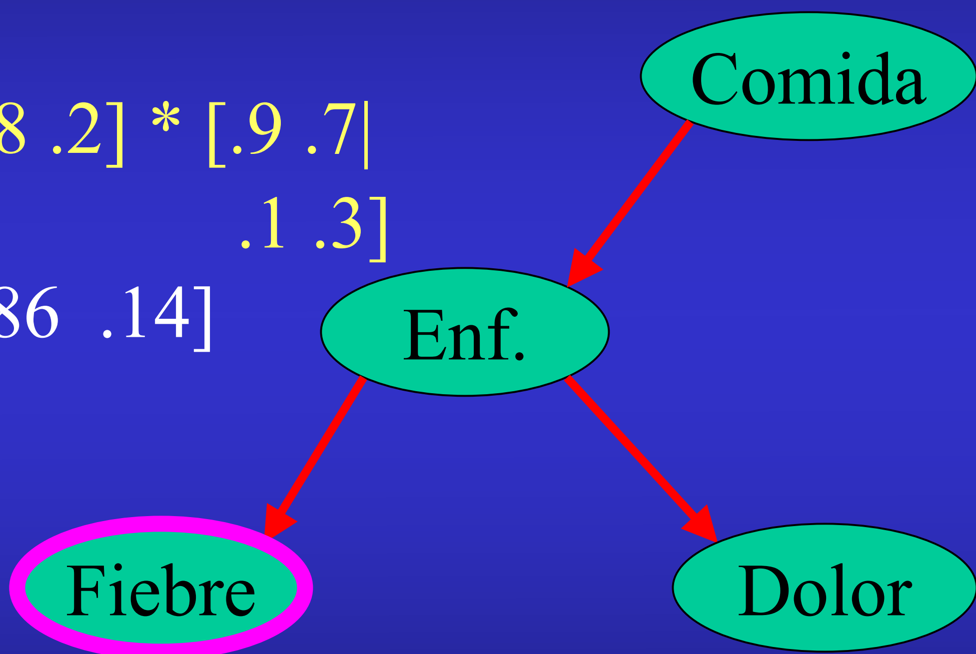
0.7 0.4

0.3 0.6

Ejemplo

$\pi(C) = [.8 \;\; .2]$

$$\pi(E) = [.8 \;\; .2] * \begin{bmatrix} .9 & .7 \\ .1 & .3 \end{bmatrix} = [.86 \;\; .14]$$



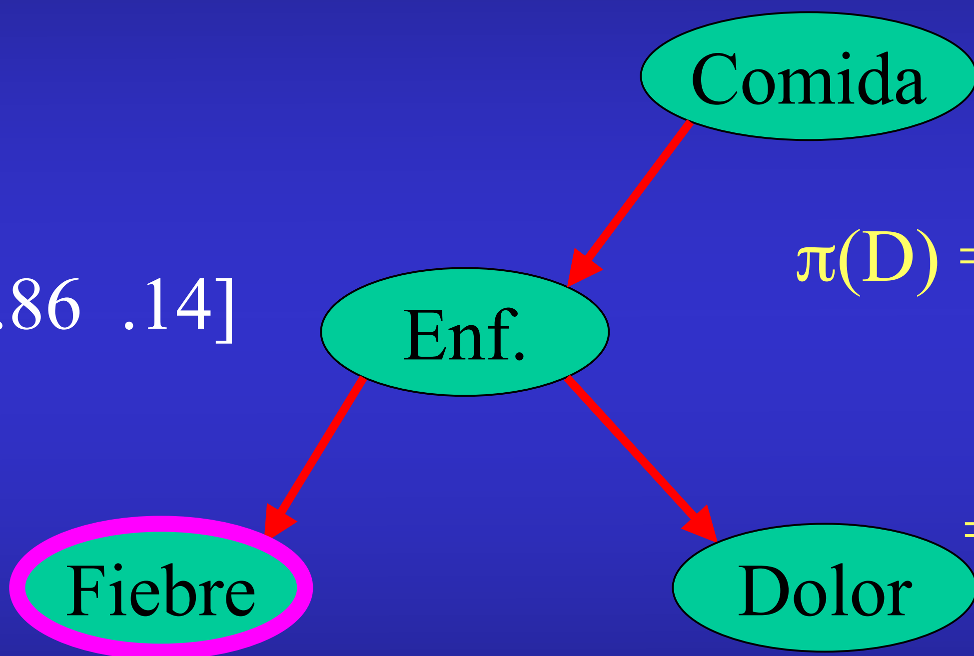
P(E|C)
0.9 0.7
0.1 0.3

P(F|E)
0.9 0.5
0.1 0.5

P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

π(C) = [.8 .2]



π(E) = [.86 .14]

π(D) = [.86 .14] * [.9 .5]
[.7 .4|
.3 .6]
= [.5698 .2742]

P(D|E)
0.7 0.4
0.3 0.6

Ejemplo

$\pi(C) = [.8 \ \ .2]$

$\lambda(C) = [.86 \ \ .78]$

$P(C) = \alpha[.688 \ \ .156]$

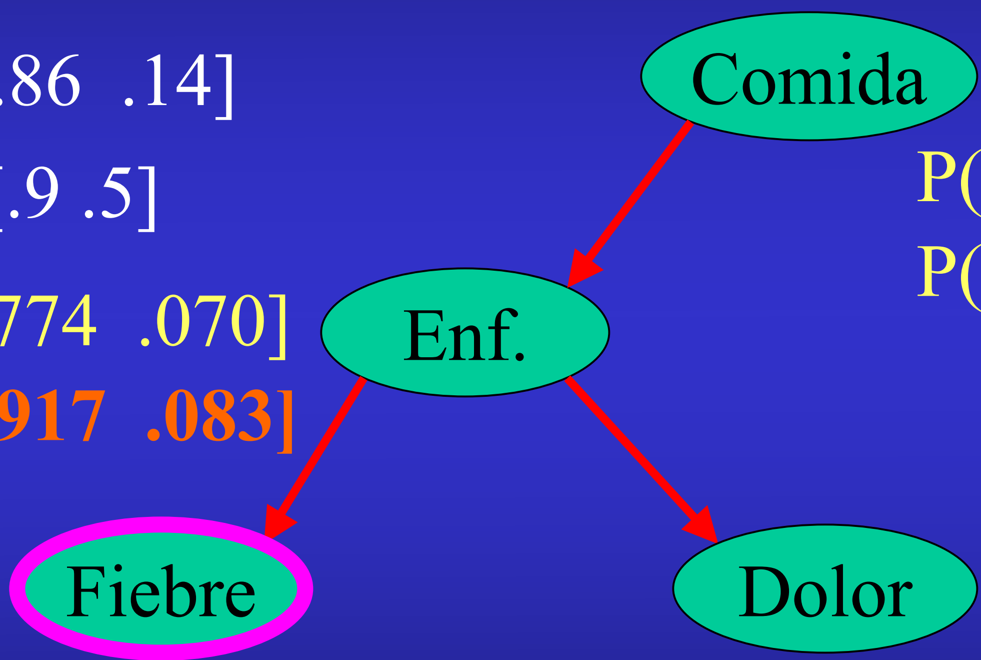
$P(C) = $ **[.815 .185]**

$\pi(E) = [.86 \ \ .14]$

$\lambda(E) = [.9 \ .5]$

$P(E) = \alpha[.774 \ \ .070]$

$P(E) = $ **[.917 .083]**



$\pi(D) = [.57 \ \ .27]$

$\lambda(D) = [1,1]$

$P(D) = \alpha[.57 \ \ .27]$

$P(D) = $ **[.67 .33]**

Tarea

- Leer sobre redes bayesianas (capítulo en la página)