

Redes Bayesianas: Introducción

- **Eliminan algunos de los problemas asociados al razonamiento probabilístico**
- **Desarrolladas a finales de los 70 (Pearl), se convirtieron durante los 90 en un esquema general de representación de la incertidumbre**
- **Una Red Bayesiana (RB) proporciona una forma compacta y modular de representar la distribución conjunta de varias variables aleatorias**
- **Una RB consta de:**
 - Una parte **cualitativa** que describe las relaciones entre las distintas variables
 - Una parte **cuantitativa** que describe la fuerza de dichas relaciones mediante probabilidades condicionadas

Redes Bayesianas: Inferencia, Decisión y Aprendizaje

- En una RB, la información proporcionada por una o más variables que se observan (evidencia) se propaga por la red y actualiza nuestra creencia acerca de las variables no observadas. A este proceso se le llama **inferencia**.
- Es posible aprender las probabilidades condicionales que describen las relaciones entre las variables a partir de los datos. Incluso es posible **aprender** la estructura completa de la red a partir de datos completos o con algunos de sus valores desconocidos.
- Las RB pueden utilizarse para tomar **decisiones** óptimas introduciendo posibles acciones y la utilidad de sus resultados

Redes Bayesianas en la Prensa

- Cnet.com

NEWS.COM SPECIAL REPORT

Old school



Probability theorist Thomas Bayes. If you've used Google, you've gone Bayesian.

18th-century theory is new force in computing

By [Michael Kanellos](#)
Staff Writer, CNET News.com
February 18, 2003, 4:00 AM PT

Thomas Bayes, one of the leading mathematical lights in computing today, differs from most of his colleagues: He has argued that the existence of God can be derived from equations. His most important paper was published by someone else. And he's been dead for 241 years.

Yet the 18th-century clergyman's theories on probability have become a major part of the mathematical foundations of application development.

Search giant [Google](#) and [Autonomy](#), a company that sells information retrieval tools, both employ Bayesian principles to provide likely (but technically never exact) results to data searches. Researchers are also using Bayesian models to determine correlations between specific symptoms and diseases, create personal robots, and develop artificially intelligent devices that "think" by doing what data and experience tell them to do.

Bayes' theorem

$$P(H|E, c) = \frac{P(H|c) P(E|H, c)}{P(E|c)}$$

Despite the esoteric symbols, the idea--roughly speaking--is simple: The likelihood that something will happen can be plausibly estimated by how often it occurred in the past. Researchers are applying the idea to everything from gene studies to filtering e-mail.

A detailed mathematical rundown can be found on the University of Minnesota's [Web site](#). And a [Bayes Rule Applet](#) on Gametheory.net lets you answer questions such as "How worried should you be if you test positive for some disease?"

"Bayesian research is used to make the best gambles on where I should flow with computation and bandwidth," said Eric Horvitz, senior researcher and [group manager of the Adaptive Systems & Interaction Group at Microsoft Research](#). "I personally believe that probability is at the foundation of any intelligence in an uncertain world where you can't know everything."

One of the more vocal Bayesian advocates is [Microsoft](#). The company is employing ideas based on probability--or "probabilistic" principles--in its Notification Platform. The technology will be embedded in future Microsoft software and is intended to let computers and cell phones automatically filter messages, schedule meetings without their owners' help and derive strategies for getting in touch with other people.

If successful, the technology will give rise to "context servers"--electronic butlers that will interpret people's daily habits and organize their lives under constantly shifting circumstances.

Developers are using Bayes' centuries-old ideas to take on the Information Age.

They're building programs designed to automatically manage the deluge of data that gets thrown at people each day via e-mail, cell phones, instant messaging programs and the like.

Such a program would collect data over time and create a model of a person's past behavior to determine how best to deal with a newly arrived message.

A simplified example:

1) E-mail arrives
From: The boss
Subject: Read me
Sent: Thu 3:10 PM

Program accesses calendar software
"Thursday, 2 PM: Meeting"

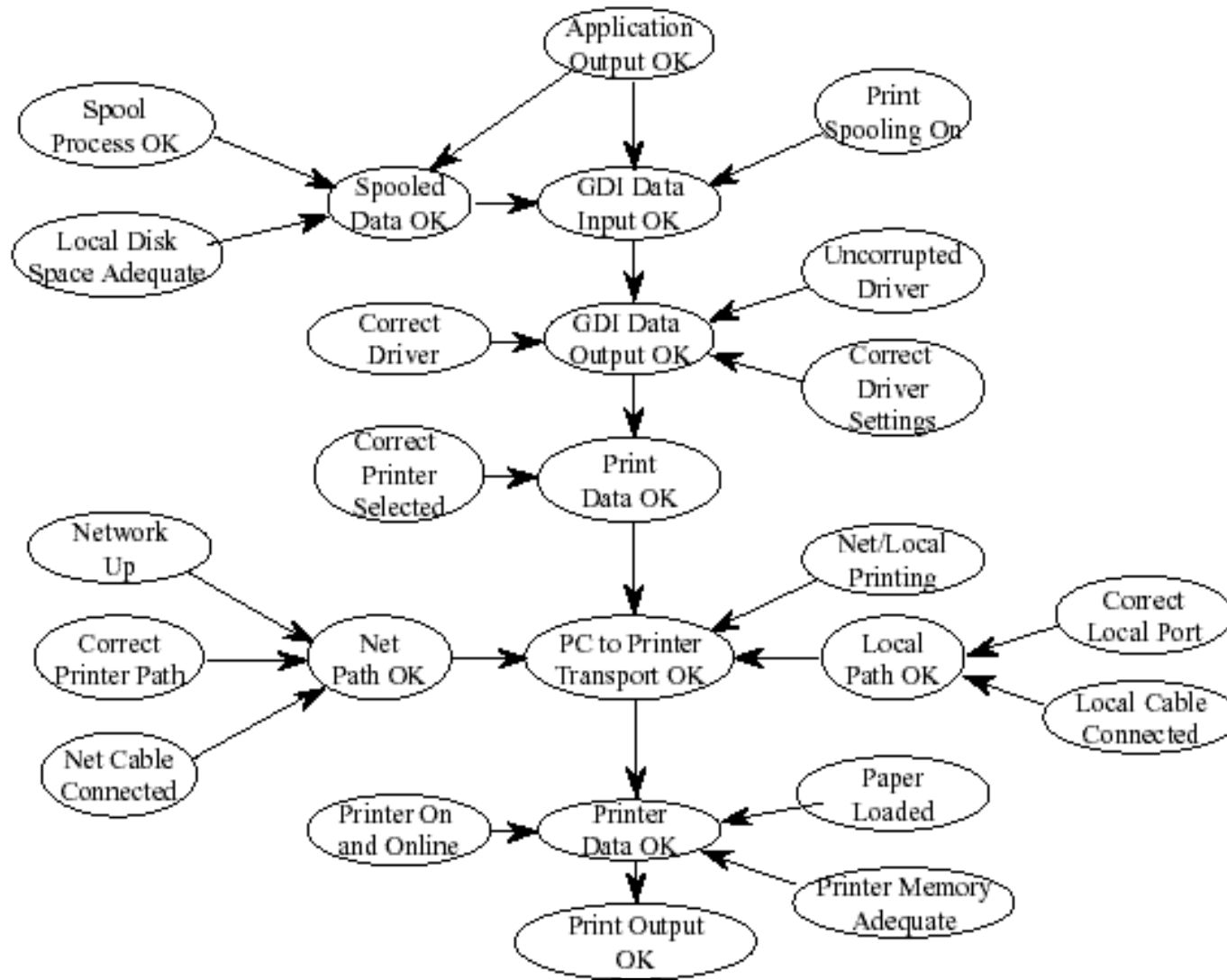
Program considers

<http://news.com.com/2009-1001-984695.html>

Redes Bayesianas: Utilización

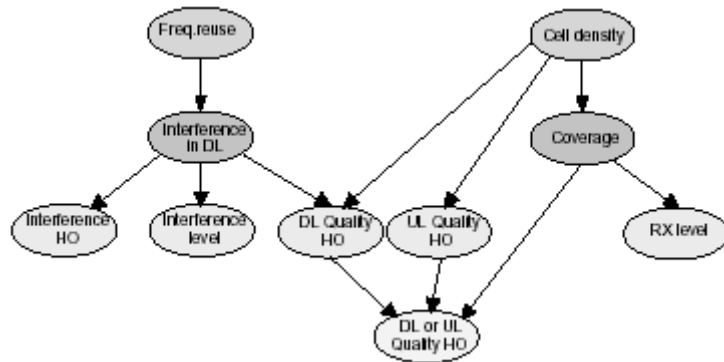
- **Algunas aplicaciones de RB en empresas**
 - Microsoft
 - Answer Wizard (Office)
 - Diagnóstico de problemas de usuario (Aladdin)
 - Home Health en la red de Microsoft (MSN)
 - Intel
 - Diagnóstico de fallos de procesadores
 - HP
 - Diagnóstico de problemas de impresora
 - Nokia
 - Diagnóstico de redes celulares
 - Nasa
 - Sistema de ayuda a la decisión en misiones espaciales

Red Bayesiana: Ejemplo



Diagnóstico de Problemas de Impresión (Heckerman)

Red Bayesiana: Ejemplo



Nokia Troubleshooting Tool v2.1.3 R1 - Connected to Local Peterborough

File Options Fault cases Cells... Results Symptoms Help

Network Analysis Conditions Scheduler Results Symptoms History Log

Fault Case Results

| BSC Name | BCF Name | BTS Name | BTS Id | Start Date | TCH Reque... | TCH Drops | TCH Drops... |
|----------|-------------|----------|--------|--------------|--------------|-----------|--------------|
| DYF5056B | dyf0053bcf1 | dyf0053b | 50 | 2001-08-0... | 1225 | 169 | 138 |

Propagate Save Results... Quick Save of Results

Expert's Opinion

Most Prob. Cause: 3b_TRX_fault_Type_C_Combiner... Probab. Certain (100%)

Comments: Other useful information

DCR_10b, dy0053b

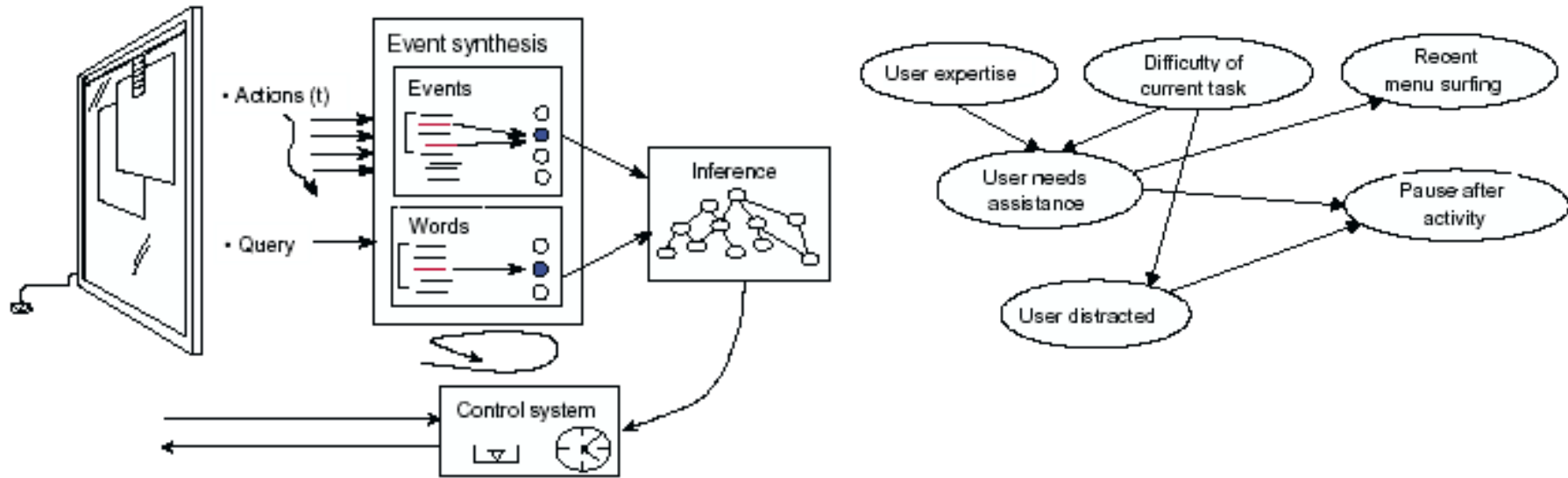
Most Prob Cause: 3b_TRX_fault_Type_C_Combiner_Non_Hopping, (99.33)

| Cause Name | Prob... | Action | Cost | Effct... | Include |
|-----------------------------|---------|---------|------|----------|---------|
| 3b_TRX_fault_Type_C_Co... | 99.33 | NO_A... | 1 | 99.33 | ✓ |
| 4_Bad_coverage_borders | 0.55 | NO_A... | 1 | 0.55 | ✓ |
| 6_Gross_antenna_misalign... | 0.1 | NO_A... | 1 | 0.1 | ✓ |
| 5_Bad_coverage_possible... | 0.01 | NO_A... | 1 | 0.01 | ✓ |
| 12_Additional_adjacencie... | 0.01 | NO_A... | 1 | 0.01 | ✓ |
| 1a_Interference_in_Cell_UL | 0 | NO_A... | 1 | 0 | ✓ |
| 1b_Interference_in_Cell_DL | 0 | NO_A... | 1 | 0 | ✓ |
| 1c_Interference_in_Cell_I01 | 0 | NO_A... | 1 | 0 | ✓ |
| 2a_Interference_in_TRX_n... | 0 | NO_A... | 1 | 0 | ✓ |
| 2b_Interference_in_TRX_n... | 0 | NO_A... | 1 | 0 | ✓ |

Analysis of cells finished. 1 FAULT CASE

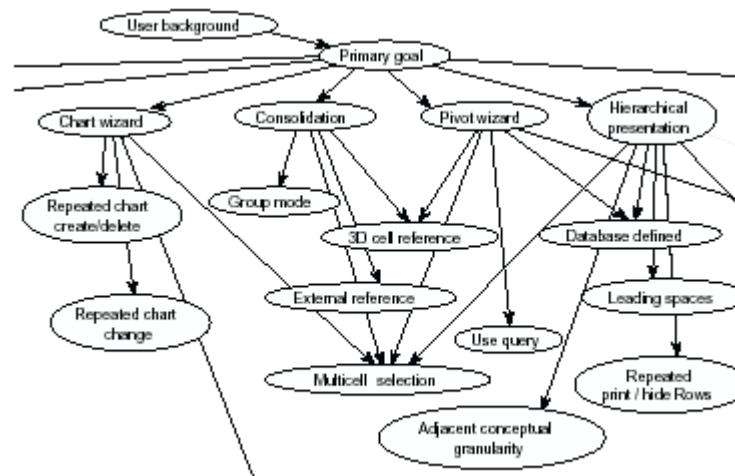
Diagnóstico de Problemas en redes celulares para Nokia (Barco y otros)

Red Bayesiana: Ejemplo



Visión de alto nivel del sistema de RB para Excel

Vista parcial de la red para inferir si el usuario tiene dificultades



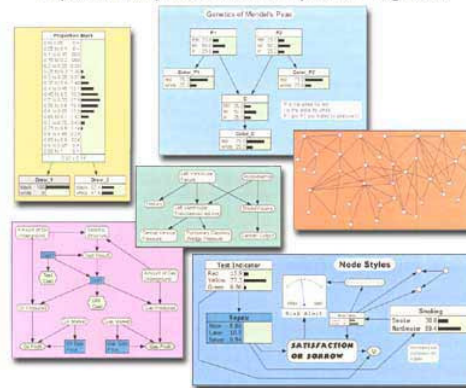
Vista parcial de la red para inferir si el usuario tiene dificultades con Excel (Heckerman)

Redes Bayesianas: Algunas Herramientas

- **Norsys**

- Programa: Netica
- Descarga de: <http://www.norsys.com/netica.html>

A Complete Software Package to Solve Problems Using Bayesian Belief Networks and Influence Diagrams



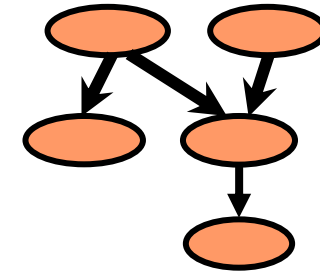
- **Microsoft**

- MSBNx
- Descarga de: <http://research.microsoft.com/adapt/MSBNx/>



¿Qué es un Red Bayesiana (RB)?

- Una RB es un **grafo dirigido** en el que **cada nodo contiene información probabilística**.
- Para determinar una RB hace falta:
 - Un **conjunto de variables aleatorias** que forman los nodos de la red. Las variables pueden ser discretas o continuas
 - Un **conjunto de enlaces dirigidos** (arcos) que conectan parejas de nodos. Si hay un arco del nodo X al Y se dice que X es el padre de Y.
 - El significado intuitivo de un arco desde el nodo X al Y es que X tiene una influencia directa en Y
 - Cada nodo X_i tiene una distribución de probabilidad condicional: $P(X_i | \text{Padres}(X_i))$ que mide el efecto de los padres de ese nodo.
 - El grafo no tiene ciclos dirigidos (y por tanto es un grafo dirigido acíclico o DAG)



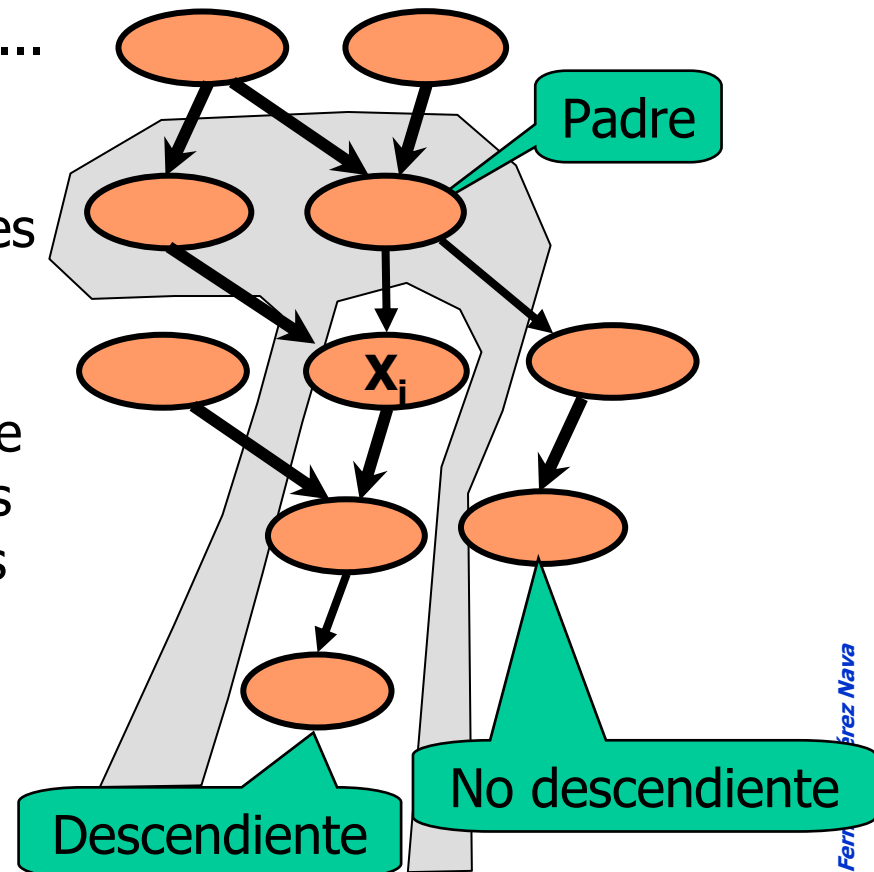
Red Bayesiana: Significado

- Los arcos en una RB proporciona una forma de codificar relaciones de independencia
- Estas relaciones se pueden especificar como:

– Dada una RB con nodos X_1, X_2, \dots, X_n . Si $\text{Padres}(X_i)$ son los padres de X_i y $\text{NoDescendientes}(X_i)$ los nodos que no son descendientes de X_i .

– Entonces para cada variable X_i se tiene que X_i es independiente de sus No Descendientes dados sus Padres. Esto lo expresamos como

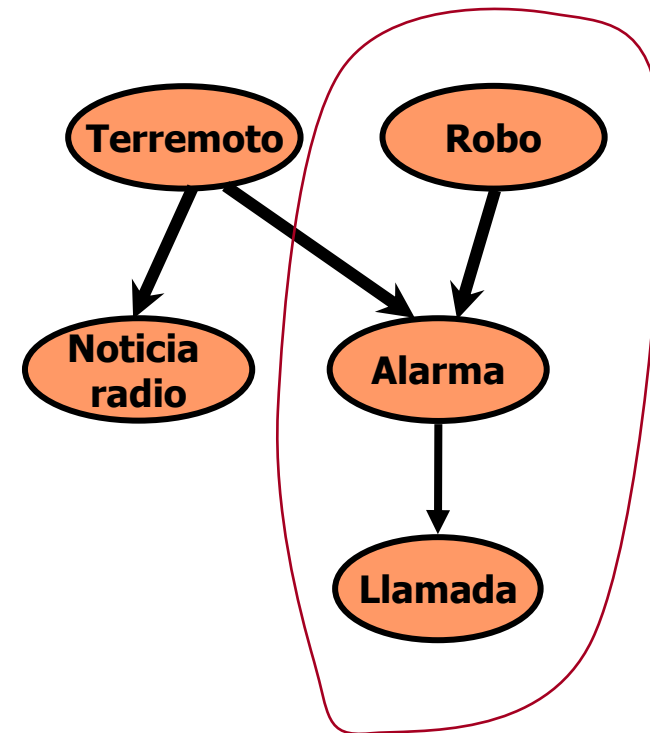
$$\text{Ind}(X_i; \text{NoDescendientes}(X_i) \mid \text{Pa}(X_i))$$



5

Ejemplos de Independencias

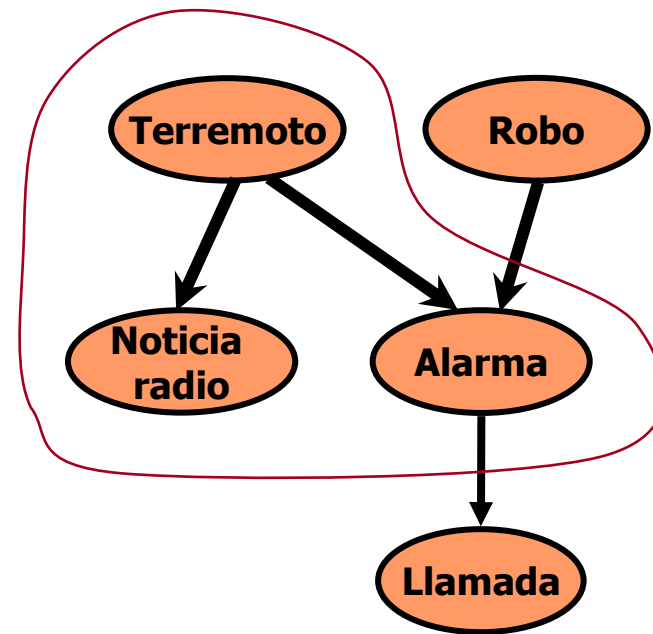
- **Para la RB del ejemplo:**
 - R y L son dependientes:
 - Si hay un robo es más probable suene la alarma, lo que hace más probable que reciba una llamada.
 - Si recibo una llamada se incrementa la probabilidad de que haya sonado la alarma y por tanto de que me hayan robado.
 - R y L son independientes si se conoce A
 - Si hay un robo ya no es más probable que suene la alarma (ya se sabe si suena o no)
 - Si recibo una llamada ya no se incrementa la probabilidad de que suene la alarma (ya se sabe si suena o no)



L es independiente de sus no-descendientes
T,R,N, dados sus padres A

Ejemplos de Independencias

- **Para la RB del ejemplo:**
 - N y A son dependientes:
 - Si oigo en la radio que ha habido un terremoto es más probable que éste haya ocurrido, lo que hace más probable que suene la alarma.
 - Si suena la alarma se incrementa la probabilidad de que haya ocurrido un terremoto y por tanto de que oiga la noticia en la radio.
 - N y A son independientes si se conoce T
 - Si oigo en la radio que ha habido un terremoto ya no es más probable que éste haya ocurrido. (ya se sabe si ha ocurrido o no).
 - Si suena la alarma ya no se incrementa la probabilidad de que haya ocurrido un terremoto (ya se sabe si ocurrió)

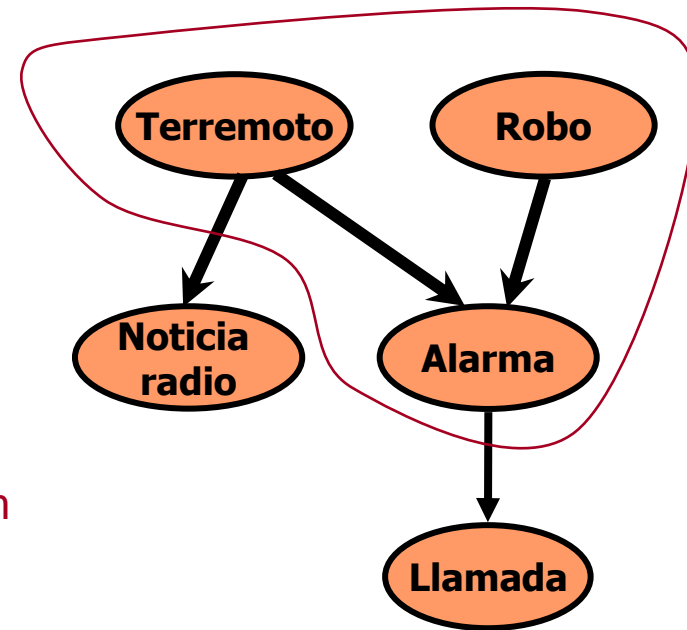


N es independiente de sus no-descendientes

R,A,L dados sus padres T

Ejemplos de Independencias

- **Para la RB del ejemplo:**
 - T y R son dependientes si se conoce A
 - Si suena la alarma y ocurre una de las causas (terremoto) me creo menos la otra (robo)
 - Si suena la alarma y ocurre una de las causas (robo) me creo menos la otra (terremoto)
 - A este efecto se le llama “**eliminación de explicaciones**”
 - T y R son independientes:
 - Si desconozco si suena la alarma y ocurre una de las causas (terremoto) no hay razón para creer menos la otra (robo)
 - Si desconozco si suena la alarma y ocurre una de las causas (robo) no hay razón para creer menos la otra (terremoto)




T es independiente de sus no-descendientes
R dados sus padres (ninguno).

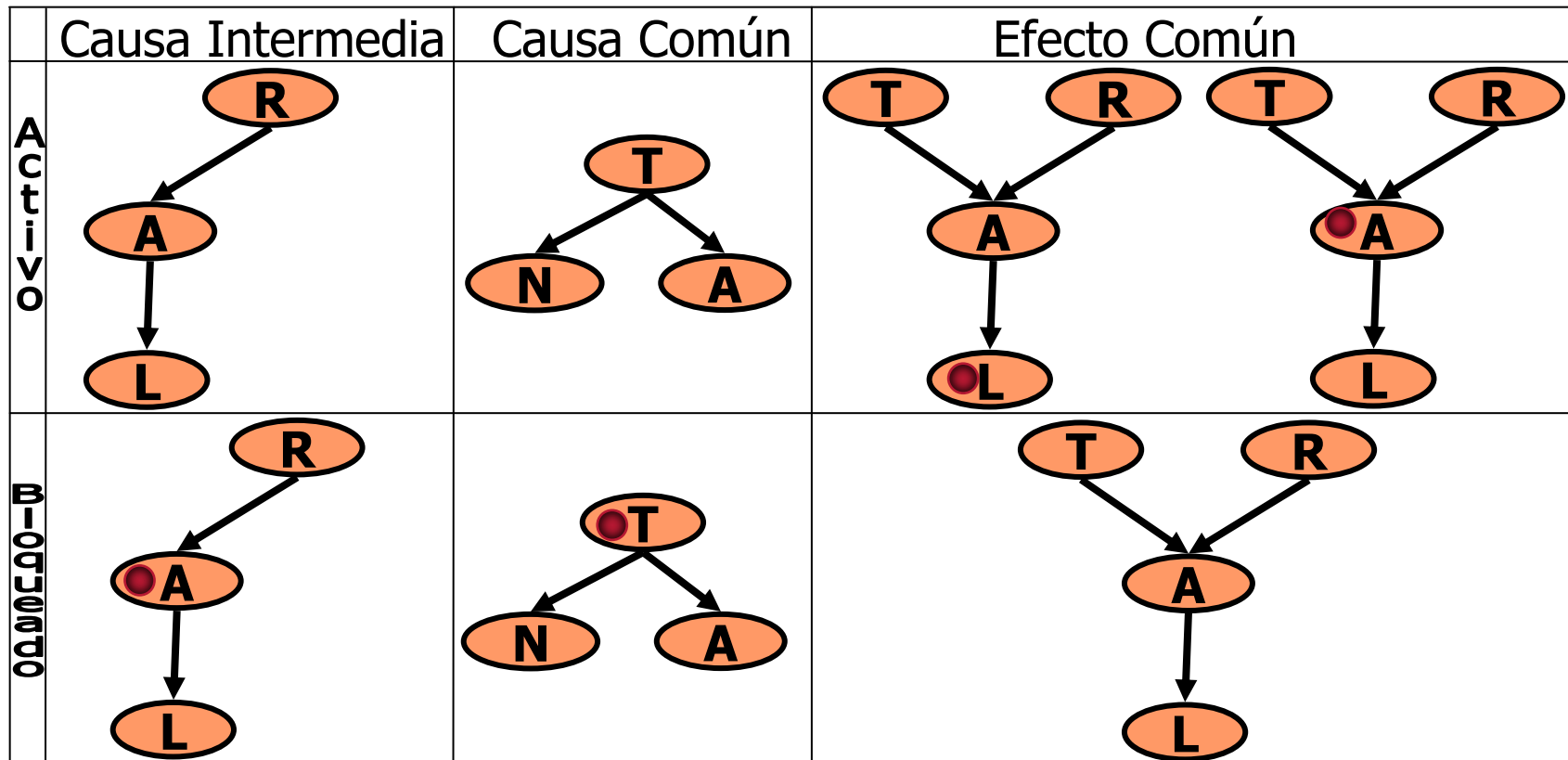
5

Transmisión de información en la red

- **Un camino del grafo puede estar:**

- Activo si pasa información por el.
- Bloqueado: si no pasa

Nodo con evidencia (observado) 



Fernando Pérez Nava

5

Teorema de Factorización

- Dada la codificación de independencias de una RB

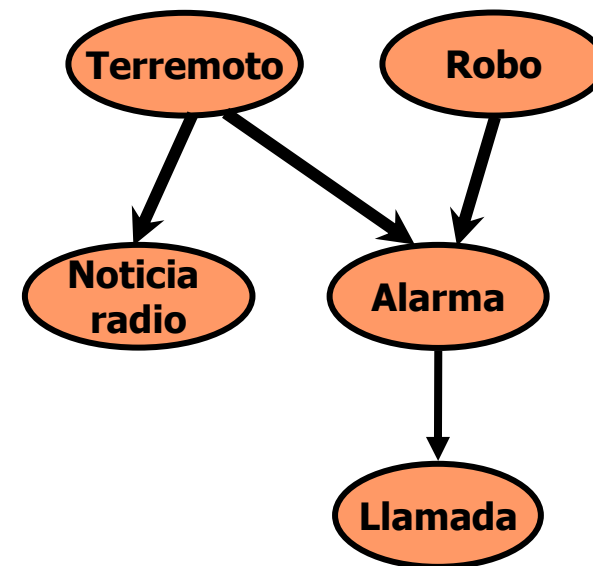
$$P(X_1, \dots, X_n) = \prod_i P(X_i \mid \text{Pa}(X_i))$$

- Ejemplo

– Teorema de Factorización:

$$P(L, A, N, T, R) =$$

$$P(R) P(T) P(N|T) P(A|R, T) P(L|A)$$



Factorización: Consecuencias

- Se puede describir P utilizando probabilidades condicionales “locales”
- Si G es un grafo disperso, es decir el número de padres de cada variable está acotado: $|\text{Pa}(X_i)| \leq k$ con k un número “pequeño” se obtiene:

- Representación compacta

El número de parámetros para describir la función de distribución conjunta es lineal en el número n de variables aleatorias $o(n)$

Nótese que el número de parámetros requerido en general es de orden $o(2^n)$

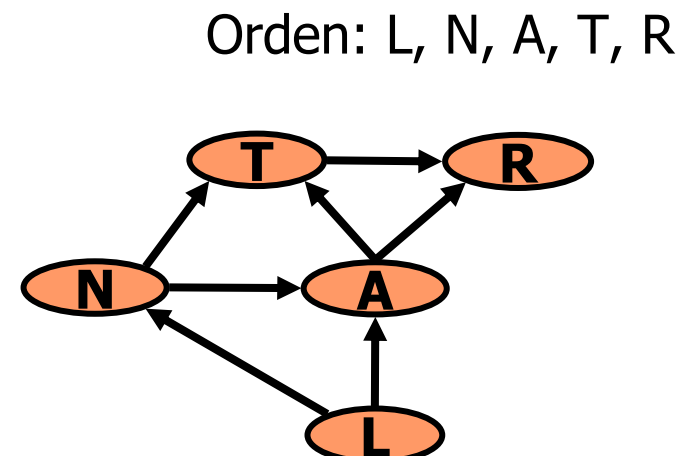
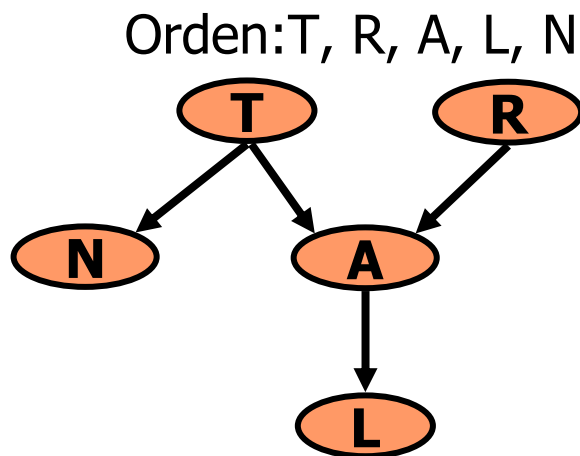
- Representación modular

Añadir una nueva variable no obliga a actualizar todos los parámetros de la representación

5

Construcción de RB

- **Un algoritmo de construcción de RB**
 - Elegir un grupo de variables X_1, \dots, X_n que describen un dominio
 - Fijar un orden en X_1, \dots, X_n (por ejemplo de las causas a los efectos)
 - Mientras haya variables
 - Elegir la siguiente variable X_i y añadir un nodo para ella
 - Seleccionar $\text{Padres}(X_i)$ como el conjunto mínimo de $\{X_1, \dots, X_{i-1}\}$, de forma que $\text{Ind}(X_i; \{X_1, \dots, X_{i-1}\} - \text{Pa}_i \mid \text{Pa}_i)$
- **La red resultante depende del orden:**



La elección de la ordenación y la causalidad

- La elección de la ordenación puede tener un **impacto drástico** en la complejidad de la Red Bayesiana.
- **Heurística para construir la RB:**
 - Construir la RB utilizando la ordenación causal entre las variables
- **Justificación**
 - Generalmente se puede asumir que los grafos generados a partir de relaciones causales cumplen las condiciones de independencia

5

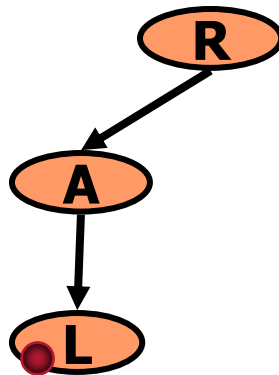
Inferencia en Redes Bayesianas

- **Inferencia:**

- Se pretende hallar la distribución de probabilidad de determinadas variables de interés dados los valores de otras variables que se observan.

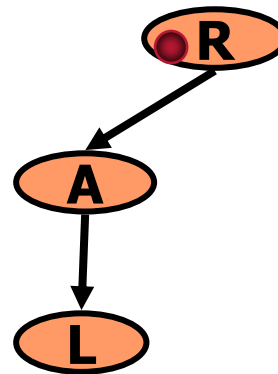
- **Principales tipos de Inferencia**

- Diagnóstico



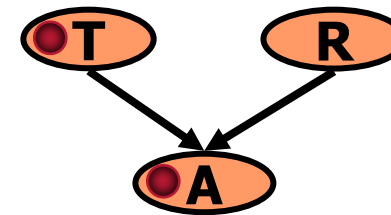
$P(\text{robo} \mid \text{llama})$

- Predicción



$P(\text{llama} \mid \text{robo})$

- Intercausal



$P(\text{robo} \mid \text{alarma}, \neg \text{terremoto})$

5

Inferencia en Redes Bayesianas

- **De manera formal**

- Supondremos que:

- La red bayesiana está formada por las variables: $\{ X_1, \dots, X_n \}$
- Las variables de interés son $\mathbf{X}_I = \{ X_1, \dots, X_i \}$
- Las variables observadas (con evidencia) son: $\mathbf{X}_O = \{ X_{i+1}, \dots, X_j \}$
 - Los valores que toman dichas variables (evidencia) son $\mathbf{e} = \{ e_{i+1}, \dots, e_j \}$
- El resto de variables son $\mathbf{X}_R = \{ X_{j+1}, \dots, X_n \}$

- El **problema a resolver** es:

Calcular:

$$P(\mathbf{X}_I | \mathbf{X}_O = \mathbf{e}) = \frac{P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e})}{P(\mathbf{X}_O = \mathbf{e})} =$$

$$\frac{P(X_1, X_2, \dots, X_i, X_{i+1} = e_{i+1}, X_{i+2} = e_{i+2}, \dots, X_j = e_j)}{P(X_{i+1} = e_{i+1}, X_{i+2} = e_{i+2}, \dots, X_j = e_j)}$$

Algoritmos de inferencia

- **Los diversos algoritmos propuestos se pueden dividir en:**
 - **Algoritmos exactos**
 - Calcular de forma exacta la inferencia solicitada.
 - La complejidad de resolver de forma exacta el problema general de inferencia en Redes Bayesianas es NP-duro.
 - Tipos
 - Para redes específicas:
 - » Árboles (Pearl), (Complejidad lineal)
 - » Poliárboles (Kim, Pearl), (Complejidad lineal)
 - Para redes generales
 - » Eliminación de Variables, Árbol de uniones (Lauritzen y Spiegelhalter)
 - **Algoritmos aproximados**
 - Se basan calcular de forma aproximada la inferencia solicitada simulando la distribución de la red bayesiana.
 - Aproximar una distribución con una tolerancia dada es también NP-duro.
 - Algunos algoritmos
 - Muestreo lógico (Henrion)
 - Ponderación de la verosimilitud (Fung y Chang)

Algoritmo de Inferencia por Eliminación de Variables

- **El problema es calcular:**

$$P(\mathbf{X}_I | \mathbf{X}_O = \mathbf{e}) = \frac{P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e})}{P(\mathbf{X}_O = \mathbf{e})}$$

- **El numerador es igual a:**

$$P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e}) = \sum_{\mathbf{X}_R} P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e}, \mathbf{X}_R) =$$

$$\sum_{X_{j+1}, X_{j+2}, \dots, X_n} P(X_1, X_2, \dots, X_i, X_{i+1} = e_{i+1}, X_{i+2} = e_{i+2}, \dots, X_j = e_j, X_{j+1}, X_{j+2}, \dots, X_n)$$

- **El denominador es igual a:**

$$P(\mathbf{X}_O = \mathbf{e}) = \sum_{\mathbf{X}_I} P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e}) =$$

$$\sum_{X_1, X_2, \dots, X_i} P(X_1, X_2, \dots, X_i, X_{i+1} = e_{i+1}, X_{i+2} = e_{i+2}, \dots, X_j = e_j)$$

- Por tanto, si se tiene el numerador el denominador se puede calcular a partir de éste.

Eliminación de Variables: Cálculo del numerador

- El problema ahora es calcular:

$$P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e}) = \sum_{\mathbf{X}_R} P(\mathbf{X}_I, \mathbf{X}_O = \mathbf{e}, \mathbf{X}_R) =$$

$$\sum_{X_{j+1}, X_{j+2}, \dots, X_n} P(X_1, X_2, \dots, X_i, X_{i+1} = e_{i+1}, X_{i+2} = e_{i+2}, \dots, X_j = e_j, X_{j+1}, X_{j+2}, \dots, X_n)$$

- **Idea General:**

– Para realizar de forma **eficiente** la suma anterior:

- Paso 1

Usar el teorema de factorización para factorizar la distribución conjunta:

$$P(X_1, X_2, \dots, X_i, X_{i+1}, X_{i+2}, \dots, X_j, X_{j+1}, X_{j+2}, \dots, X_n) = \prod_i P(X_i, \text{pa}(X_i))$$

- Paso 2

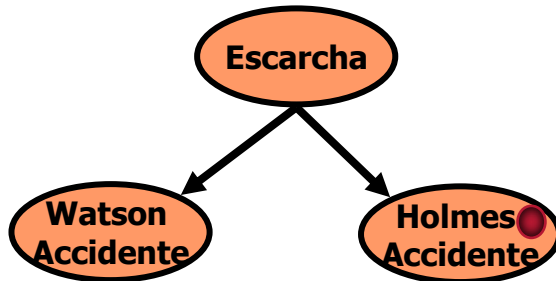
Fijar las variables observadas a sus valores de evidencia

- Paso 3

Eliminar de forma iterativa del sumatorio las variables que no son de interés ni de evidencia

Eliminación de variables: Ejemplo con evidencia

- La red de la Escarcha



| P(E) | | P(W E) | | P(H E) | |
|------|----|--------|-----|--------|-----|
| | | W | ¬W | H | ¬H |
| E | e | 0.7 | 0.8 | 0.8 | 0.2 |
| | ¬e | 0.3 | 0.1 | 0.1 | 0.9 |

- Calculemos $P(w | h)$

– Factorización:

$$P(E)P(W|E)P(H|E)$$

Sustitución de la evidencia

$$P(E)P(W|E)P(h|E)$$

Eliminamos E

$$P(W, h)$$

Normalizamos

$$P(W|h)$$

| $g_E(W)$ | |
|----------|---|
| w | $0.7 \times 0.8 \times 0.8 + 0.3 \times 0.1 \times 0.1 = 0.451$ |
| ¬w | $0.7 \times 0.2 \times 0.8 + 0.3 \times 0.9 \times 0.1 = 0.139$ |

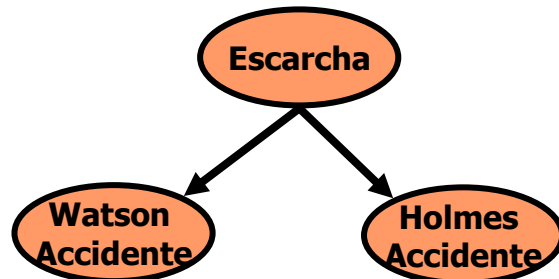
$$g_E(W) = \sum_E P(E)P(W|E)P(h|E)$$

Normalizamos

| | |
|----|----------------|
| w | ≈ 0.77 |
| ¬w | ≈ 0.23 |

Eliminación de variables: Ejemplo sin evidencia

- La red de la Escarcha



- Calculemos $P(W)$

– Factorización:

$$P(E)P(W|E)P(H|E)$$

Eliminamos H

$$g_H(E) = \sum_H P(H|E)$$

$$P(E)P(W|E)g_H(E)$$

Eliminamos E

$$g_E(W) = \sum_E g_H(E)P(E)P(W|E)$$

$$P(W)$$

| P(E) | | P(W E) | | P(H E) | |
|------|-----|--------|-----|--------|-----|
| E | | W | | H | |
| | | w | ¬w | h | ¬h |
| e | 0.7 | 0.8 | 0.2 | 0.8 | 0.2 |
| ¬e | 0.3 | 0.1 | 0.9 | 0.1 | 0.9 |

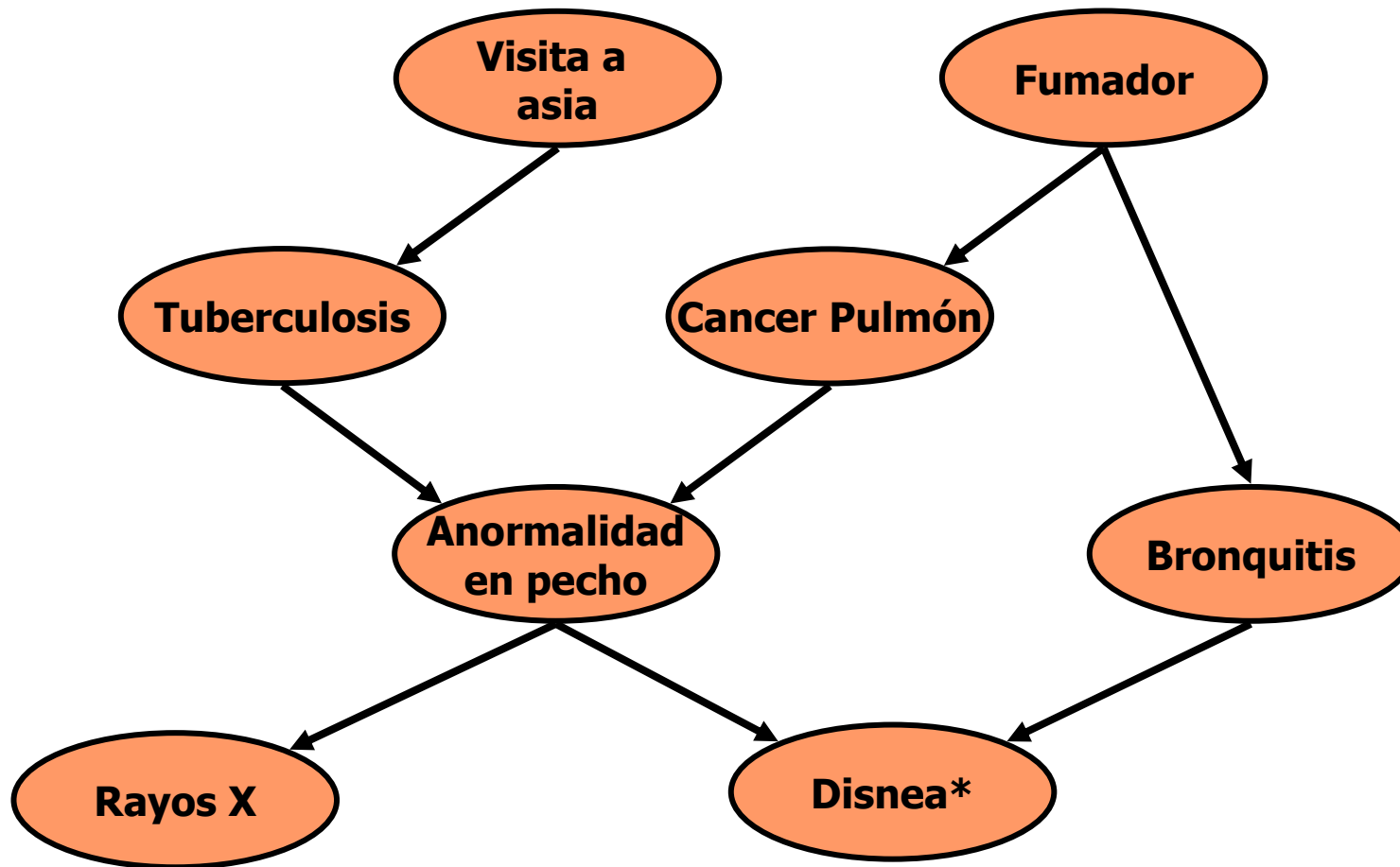
| g _H (E) | |
|--------------------|-----------------|
| E | |
| | |
| e | 0.8 + 0.2 = 1.0 |
| ¬e | 0.1 + 0.9 = 1.0 |

| | w | ¬w |
|----|--|--|
| w | 1.0 × 0.7 × 0.8 + 1.0 × 0.3 × 0.1 = 0.59 | |
| ¬w | | 1.0 × 0.7 × 0.2 + 1.0 × 0.3 × 0.9 = 0.41 |

5

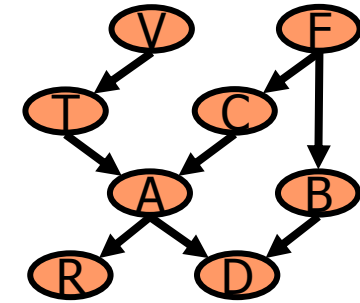
Un Ejemplo más Complejo

- La Red Asia



*Disnea=Dificultad para respirar

Eliminación de variables: Ejemplo con evidencia



- **Ejemplo: calcular $P(C | v, f, d)$**

- Como siempre calculamos $P(C, v, f, d)$ y normalizamos

- Escribimos la factorización:

$$P(V)P(F)P(T | V)P(C | F)P(B | F)P(A | T, C)P(R | A)P(D | A, B)$$

y sustituimos la evidencia $V=\text{cierto}, F=\text{cierto}, D=\text{cierto}$

$$P(v)P(f)P(T | v)P(C | f)P(B | f)P(A | T, C)P(R | A)P(d | A, B)$$

tenemos por tanto que eliminar: R, T, A, B (puesto que V, F, D están fijos)

- Proceso de eliminación:

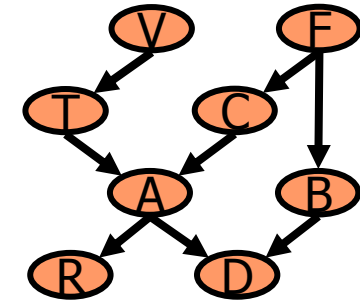
$$P(v)P(f)P(T | v)P(C | f)P(B | f)P(A | T, C)P(R | A)P(d | A, B)$$

Eliminamos R

$$h_R(A) = \sum_R P(R | A)$$

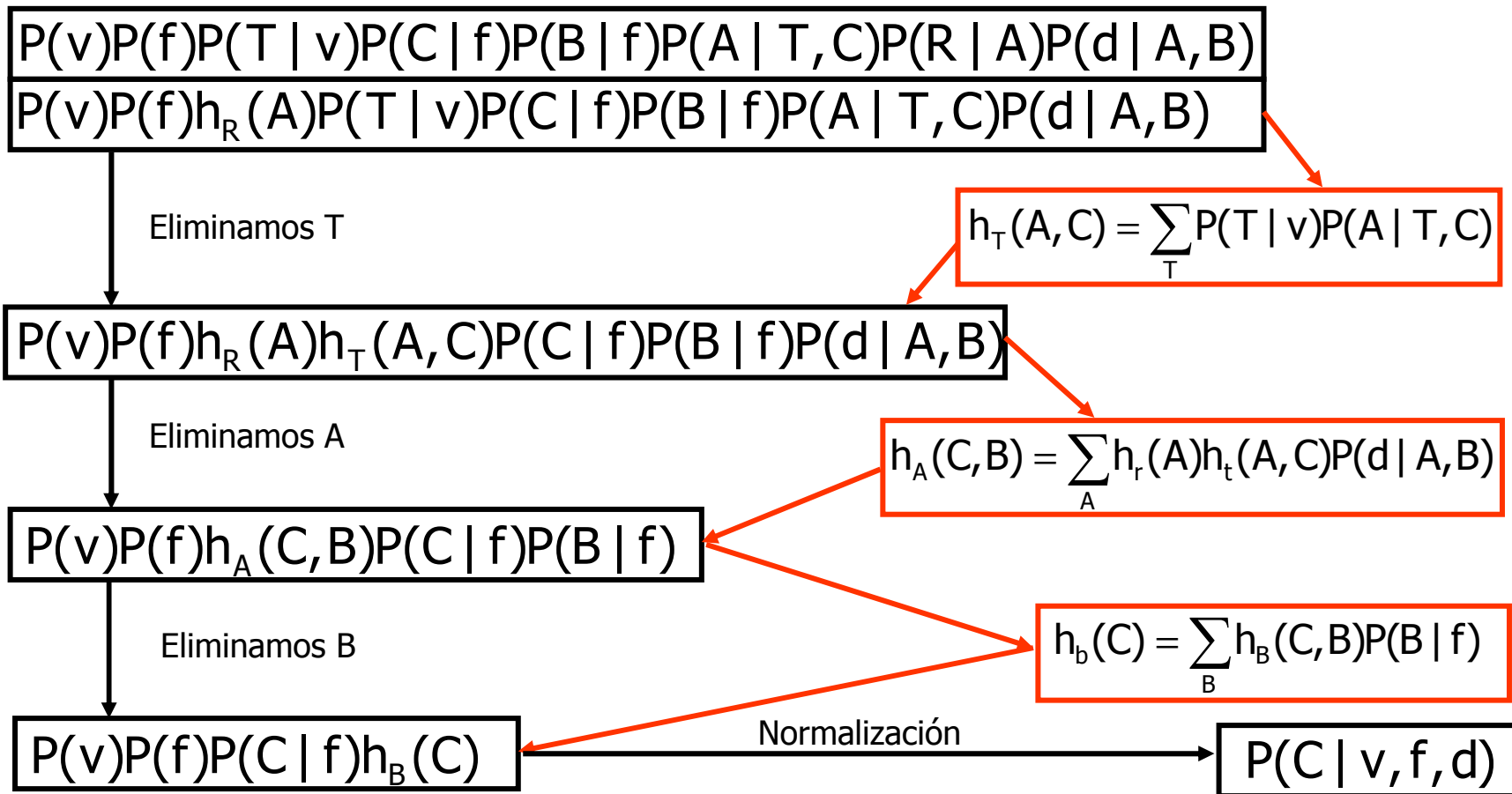
$$P(v)P(f)h_R(A)P(T | v)P(C | f)P(B | f)P(A | T, C)P(d | A, B)$$

Eliminación de variables: Ejemplo con evidencia

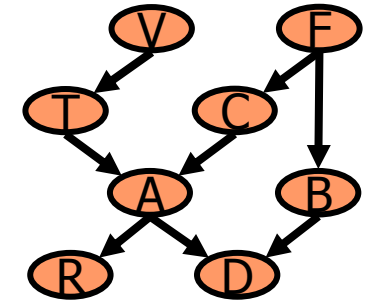


Proceso de eliminación: continuación

– Eliminando: T,A,B



Eliminación de variables: Ejemplo sin evidencia



- **Ejemplo: Calcular $P(D)$**
 - Eliminaremos: V, F, R, T, C, A, B
 - Aplicando el teorema de factorización:

$$P(V)P(F)P(T | V)P(C | F)P(B | F)P(A | T, C)P(R | A)P(D | A, B)$$

Eliminamos V

$$h_V(T) = \sum_V P(T | V)P(V)$$

$$h_V(T)P(F)P(C | F)P(B | F)P(A | T, C)P(R | A)P(D | A, B)$$

Eliminamos F

$$h_F(C, B) = \sum_F P(C | F)P(B | F)P(F)$$

$$h_V(T)h_F(C, B)P(A | T, C)P(R | A)P(D | A, B)$$

Eliminamos R

$$h_R(A) = \sum_A P(R | A)$$

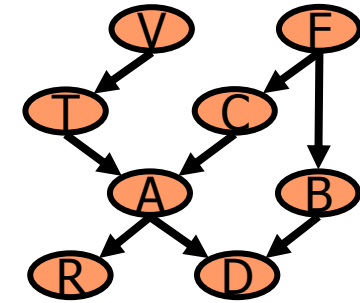
$$h_V(T)h_F(C, B)h_R(A)P(A | T, C)P(D | A, B)$$

Eliminamos T

$$h_T(A, C) = \sum_T h_V(T)P(A | T, C)$$

$$h_T(A, C)h_F(C, B)h_R(A)P(D | A, B)$$

Eliminación de variables: Ejemplo con evidencia



- **Ejemplo: Calcular $P(D)$**
 - Eliminaremos V, F, R, T, C, A, B
 - Proceso de eliminación (continuación)

$$P(V)P(F)P(T | V)P(C | F)P(B | F)P(A | T, C)P(R | A)P(D | A, B)$$

$$h_V(T)P(F)P(C | F)P(B | F)P(A | T, C)P(R | A)P(D | A, B)$$

$$h_V(T)h_F(C, B)P(A | T, C)P(R | A)P(D | A, B)$$

$$h_V(T)h_F(C, B)h_R(A)P(A | T, C)P(D | A, B)$$

$$h_T(A, C)h_F(C, B)h_R(A)P(D | A, B)$$

Eliminamos C

$$h_C(A, B)h_R(A)P(D | A, B)$$

Eliminamos A

$$h_A(D, B)$$

Eliminamos B

$$P(D)$$

$$h_C(A, B) = \sum_C h_T(A, C)h_F(C, B)$$

$$h_A(D, B) = \sum_A h_C(A, B)h_R(A)P(D | A, B)$$

$$P(D) = \sum_B h_A(B, D)$$

Complejidad del Algoritmo de Eliminación

- **Estudiamos la complejidad:**

- En cada paso se calcula:

$$h_X(Y_1, \dots, Y_k) = \sum_x h'_X(X, Y_1, \dots, Y_k), \quad h'_X(X, Y_1, \dots, Y_k) = \prod_{i=1}^m f_i(X, Y_{i1}, \dots, Y_{ik})$$

- Entonces:

- Para cada valor de Y_1, Y_2, \dots, Y_k hacemos $|\text{Val}(X)|$ sumas, por tanto el número total de sumas es $|\text{Val}(X)| \times |\text{Val}(Y_1)| \times \dots \times |\text{Val}(Y_k)|$
- Para cada valor de X, Y_1, Y_2, \dots, Y_k hacemos m multiplicaciones, por tanto el número total de de multiplicaciones es de $m |\text{Val}(X)| \times |\text{Val}(Y_1)| \times \dots \times |\text{Val}(Y_k)|$
- Por tanto: la complejidad es **exponencial** en el número de variables de los factores intermedios
- Es necesario **buscar "buenos" ordenes de eliminación** de variables para reducir el tamaño de los factores intermedios.
- Sin embargo, el problema de buscar la ordenación óptima tiene también complejidad exponencial, por lo que se emplean diversas **heurísticas**.

Resumen

- Las redes bayesianas permiten **modelar la incertidumbre de forma probabilística** incluso cuando el número de variables es elevado.
- Se basan en utilizar la noción de **independencia** para generar una representación **compacta y modular** de la distribución probabilística de las variables.
- La tarea fundamental en Redes Bayesianas es la **inferencia**: conocer la distribución de unas variables cuando se observan otras.
- El problema general de inferencia en Redes Bayesianas es **NP-duro**.
- Existen diversos algoritmos para resolver el problema de inferencia: **exactos y aproximados**.