Modelos Gráficos Probabilistas

L. Enrique Sucar

INAOE

Sesión 8: Redes Bayesianas - Representación

"La probabilidad no es realmente sobre números, es sobre la estructura del razonamiento"

[G. Shafer]

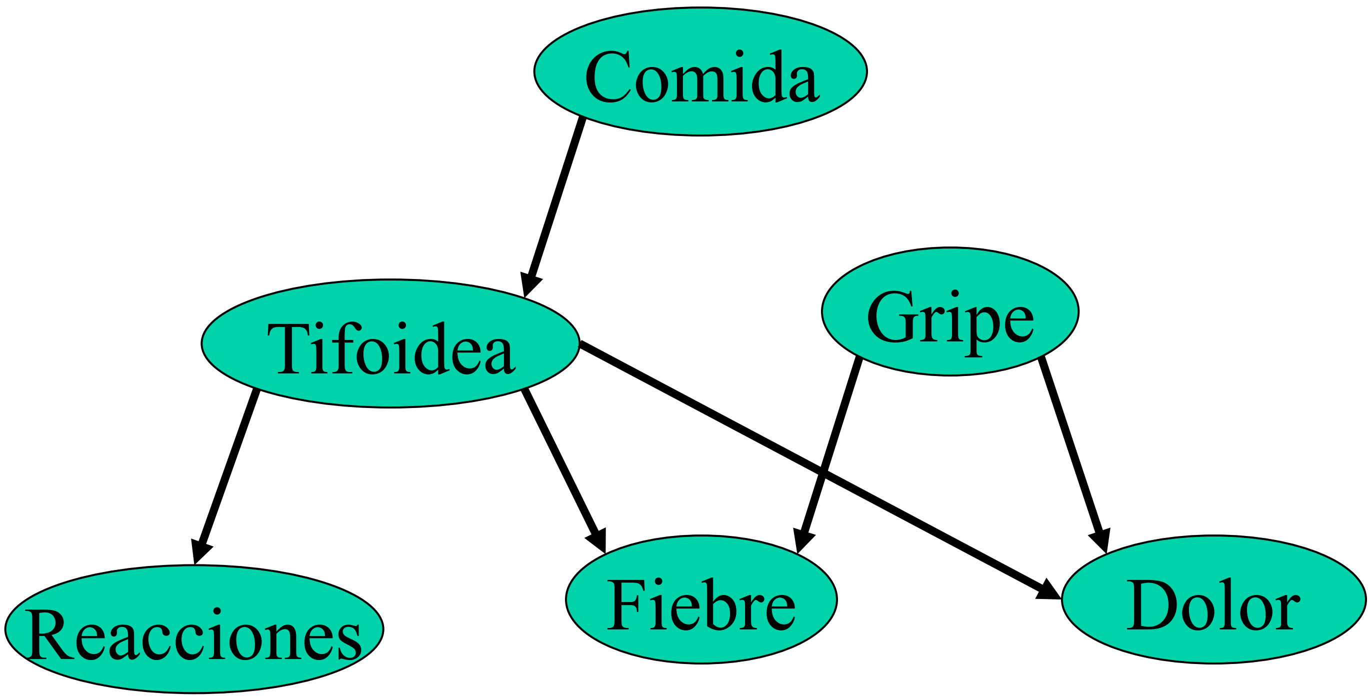
Redes Bayesianas

- Introducción
- Representación estructural
  - Separación – D
  - Correspondencia estructura – distribución de probabilidad (Mapa D, I, P)
  - Axiomas de independencia
- Representación paramétrica
  - Parámetros
  - Modelos canónicos
  - Otras representaciones

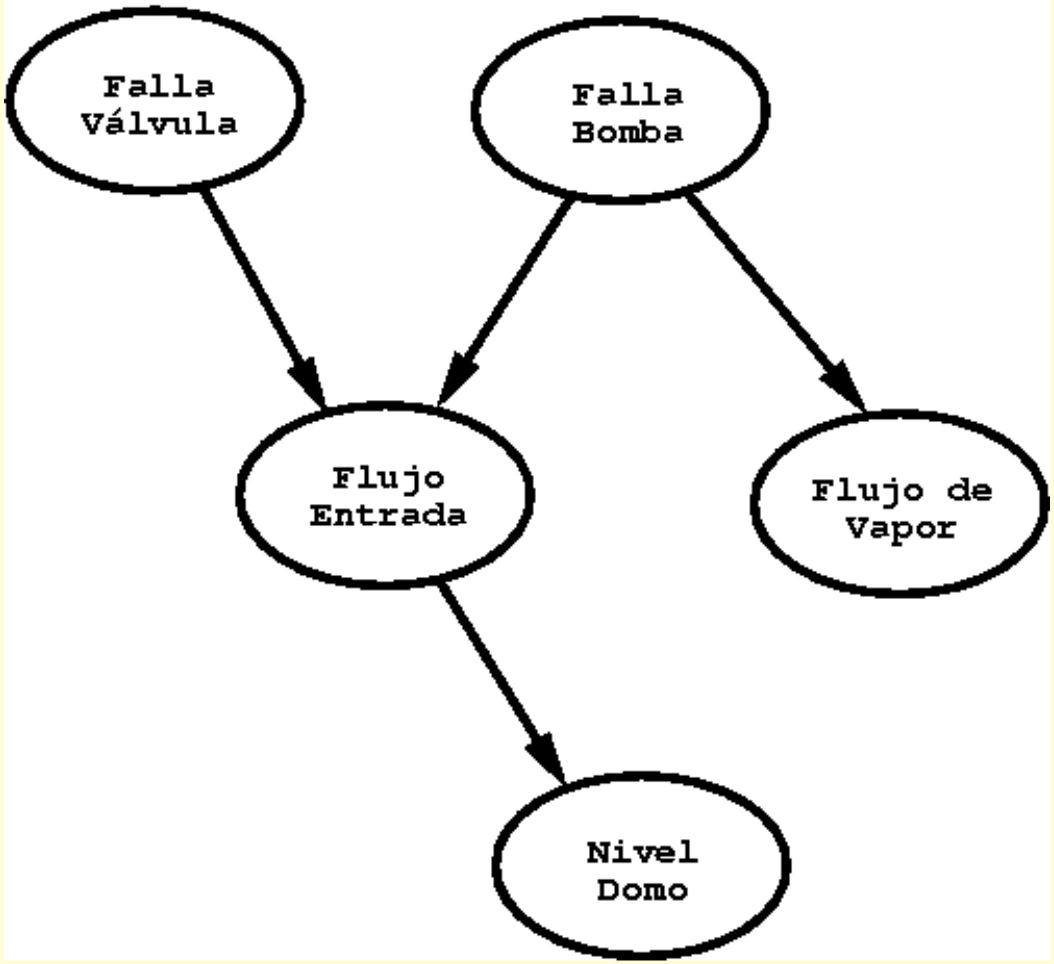
Representación

- 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



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:

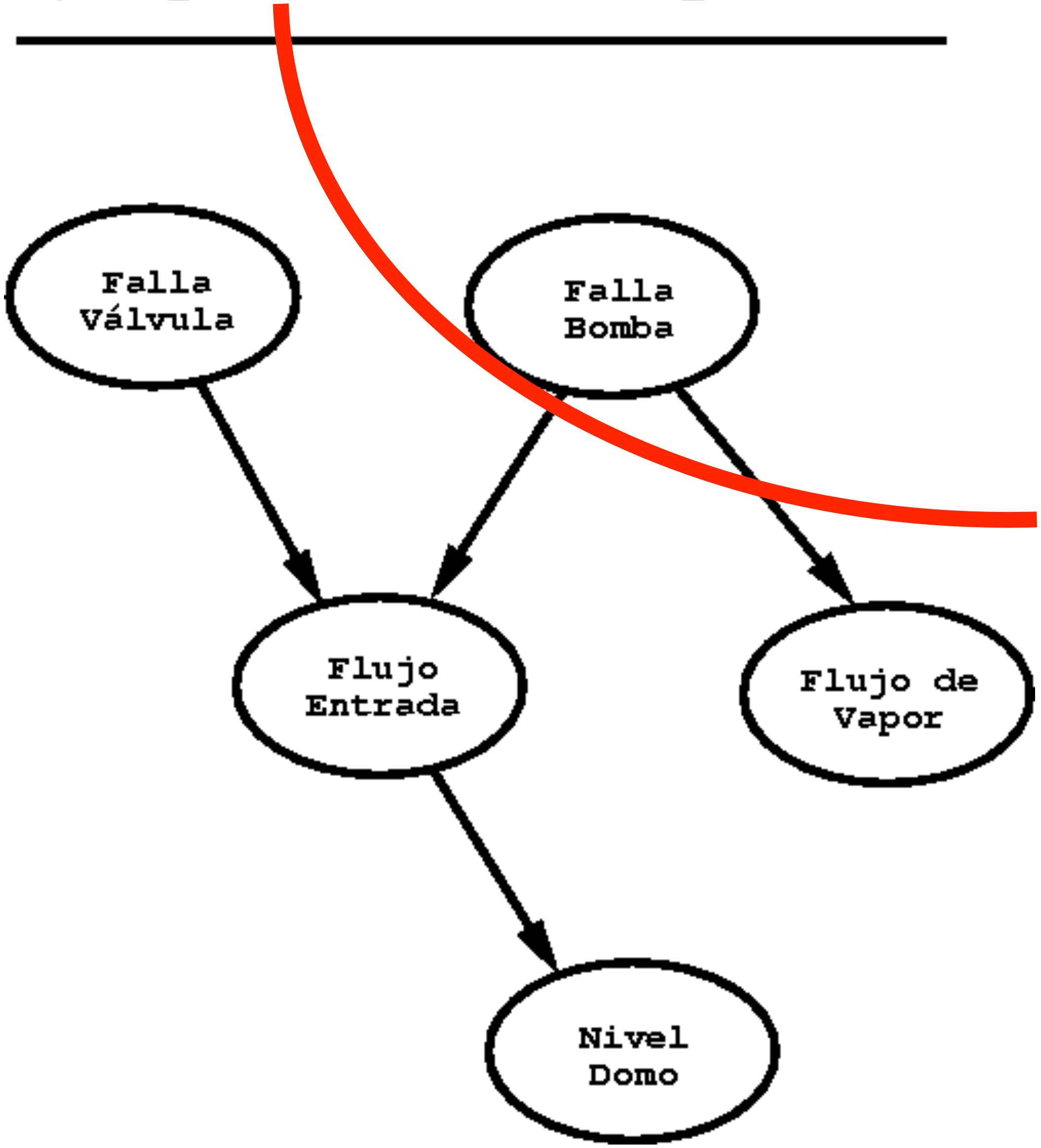
{Fva} es cond. indep. de {Fv, Fe, Nd} dado {Fb}

- 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, las 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).

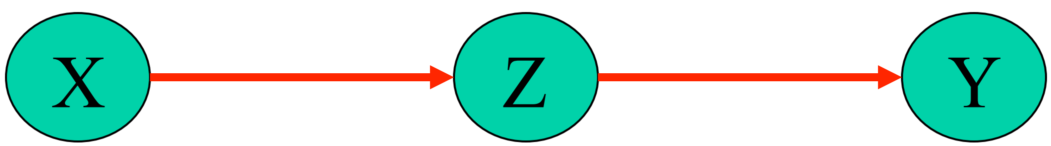
Representación Gráfica

- 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"

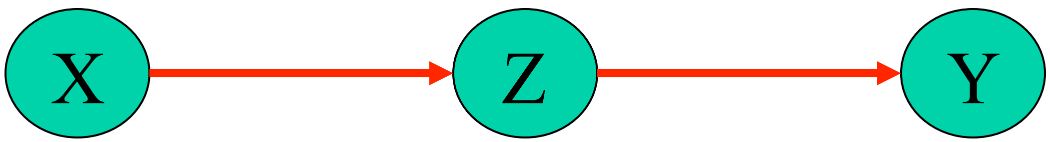
- El conjunto de variables A es independiente del conjunto B dado el conjunto C, si no existe trayectoria entre A y B en que
  1. Todos los nodos convergentes están o tienen descendientes en C
  2. Todos los demás nodos están fuera de C

Separación “D”

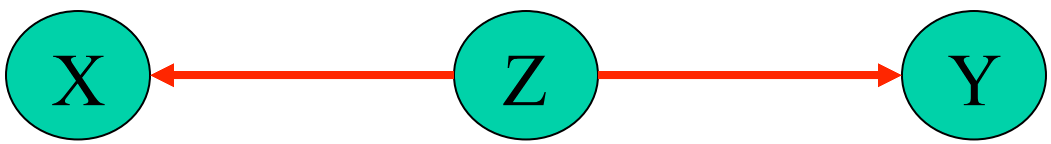
- Tres casos básicos
  - Arcos divergentes
  - Arcos en secuencia
  - Arcos convergentes

Separación "D" – casos básicos

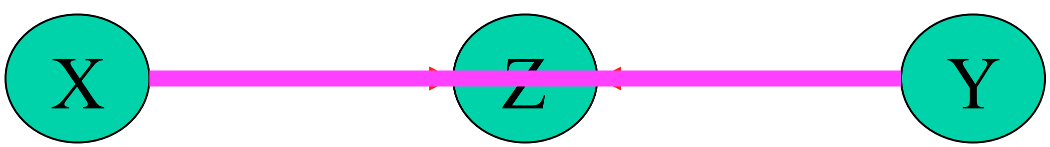
- caso 1: Secuencia:



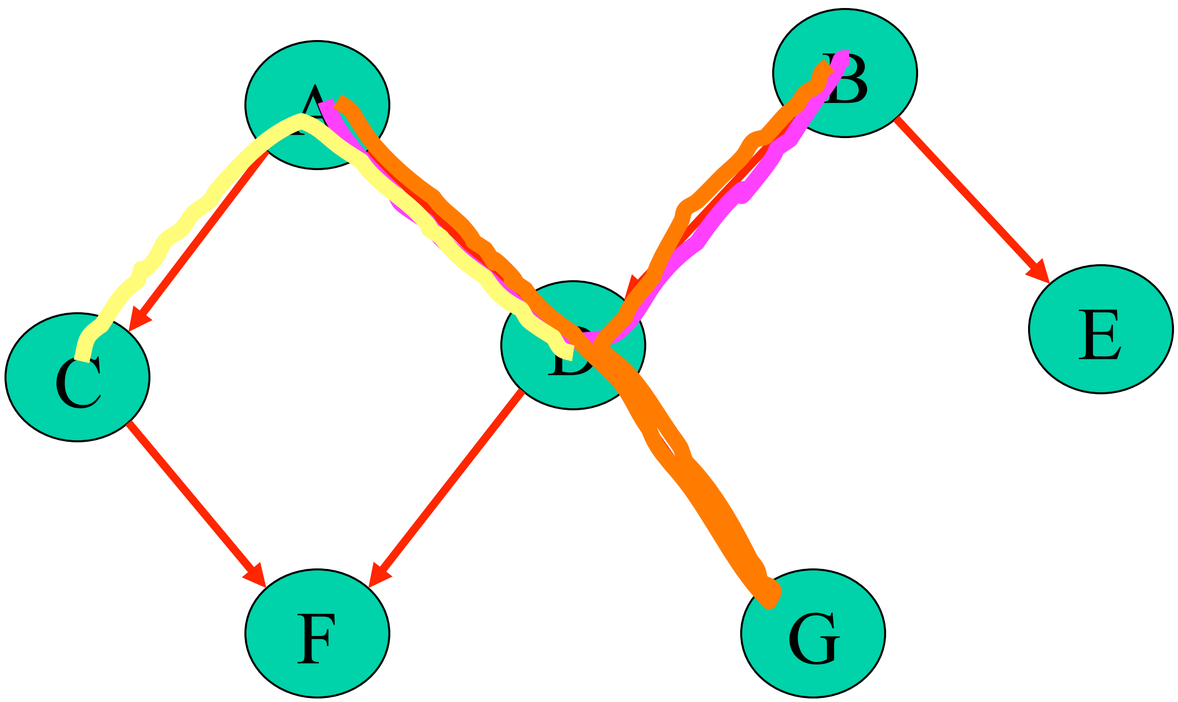
- caso 2: Divergentes:



- 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)?

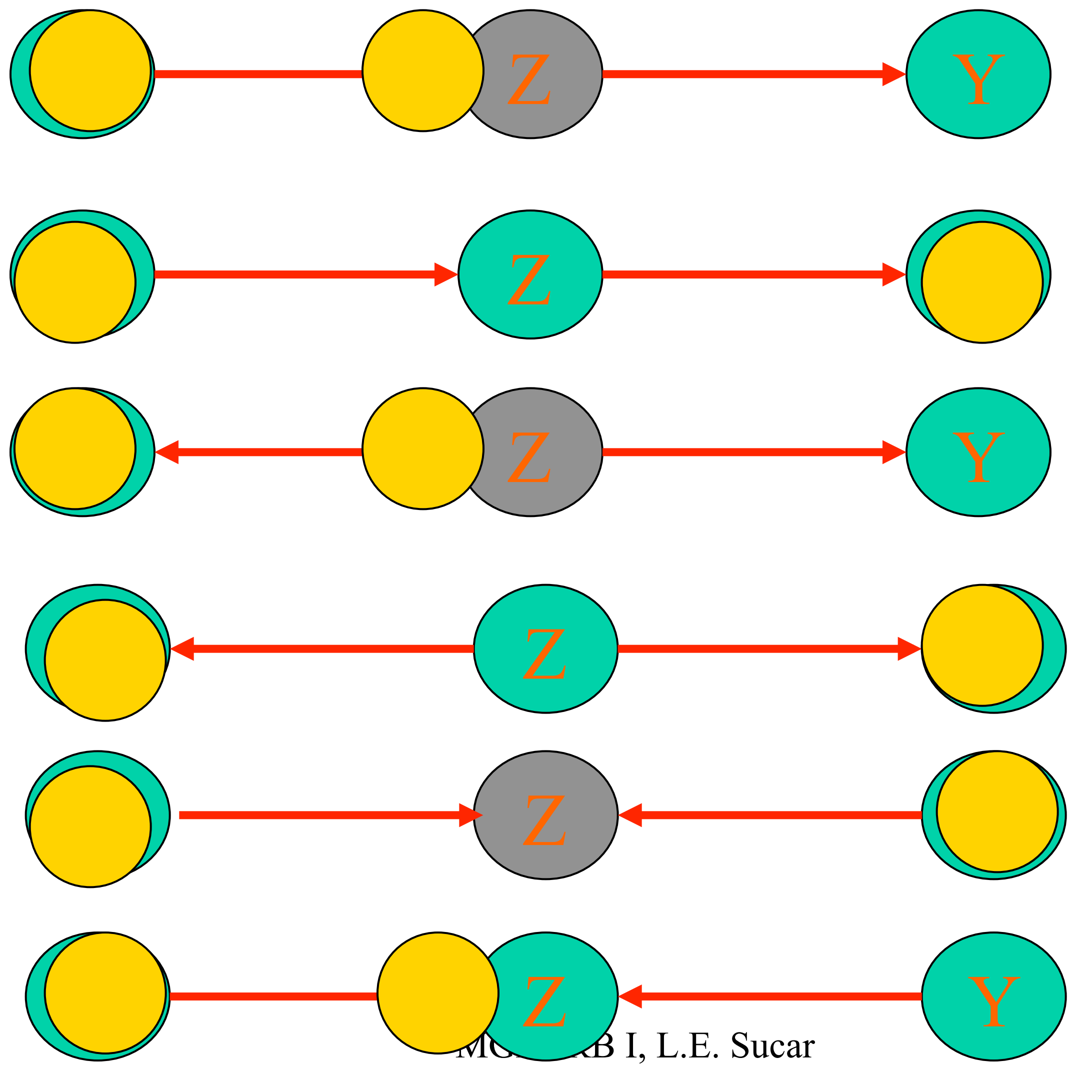
Bayes Ball

- Otra forma de ver si dos conjuntos de variables (X, Z) están separados por otro (Y) es mediante al algoritmo de la “pelota de Bayes”
- Se sombrean los nodos en Y y se lanzan pelotas desde todos los nodos en X hacia Z
- Si alguna pelota llega a Z, no son condicionalmente independientes I (X,Y,Z)

Bayes Ball

- Las reglas para el paso de las pelotas son:
  - Si el nodo es divergente (casos 1 y 2) y no está sombreado, la pelota pasa
  - Si el nodo es divergente (casos 1 y 2) y está sombreado, la pelota no pasa
  - Si el nodo es convergente (caso 3) y no está sombreado, la pelota no pasa
  - Si el nodo es convergente (caso 3) y está sombreado, la pelota pasa

Bayes ball



Correspondencia Grafo-Modelo

- Dada una distribución de probabilidad o modelo (M) y una representación gráfica de dependencias o grafo (G) debe existir una correspondencia entre las independencias representados en ambos
- Tres tipos básicos - *mapas*

Correspondencia Grafo-Modelo

- Mapa-D: las variables independientes están separadas en el grafo
- Mapa-I: las variables separadas en el grafo son independientes
- Mapa perfecto: mapa-I & mapa-D

- No es siempre posible tener un mapa perfecto (hay distribuciones con relaciones de independencia que no se pueden representar como un GAD)

Correspondencia Grafo-Modelo

- Mapa-I mínimo: las variables separadas en el grafo son independientes y al quitar cualquier arco se destruye esta condición

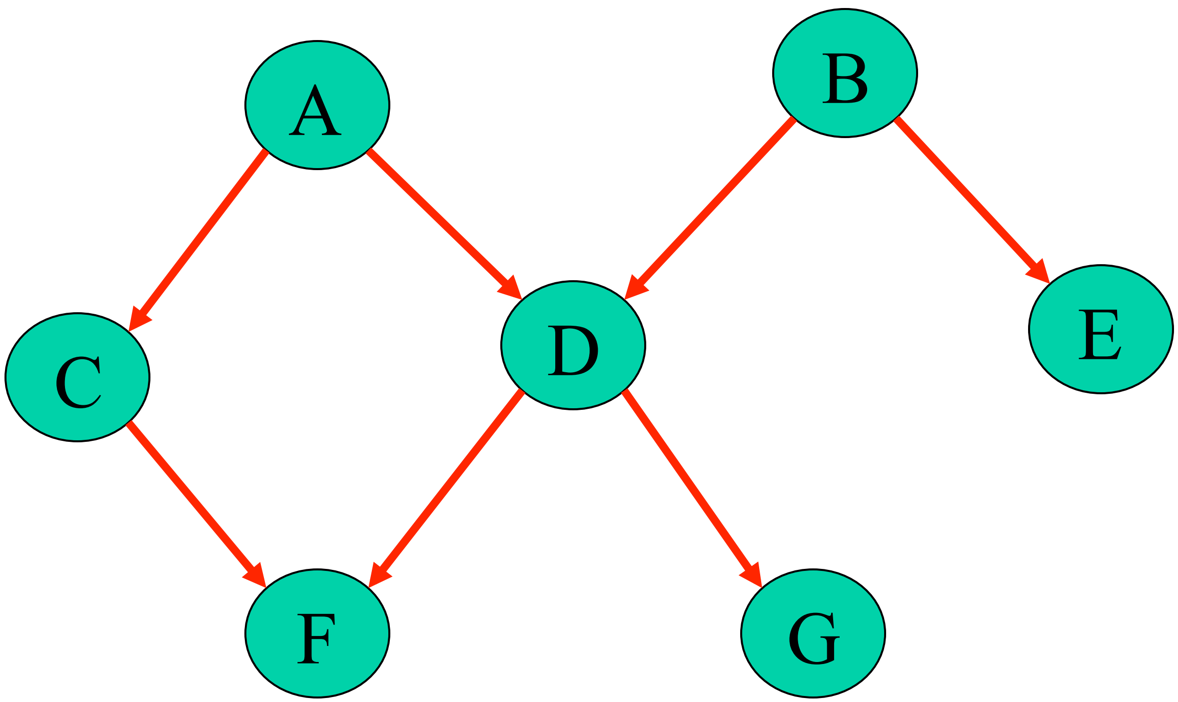
- Una red bayesiana es un grafo acíclico dirigido (GAD) que corresponde a un mapa-I mínimo de una distribución de probabilidad P

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

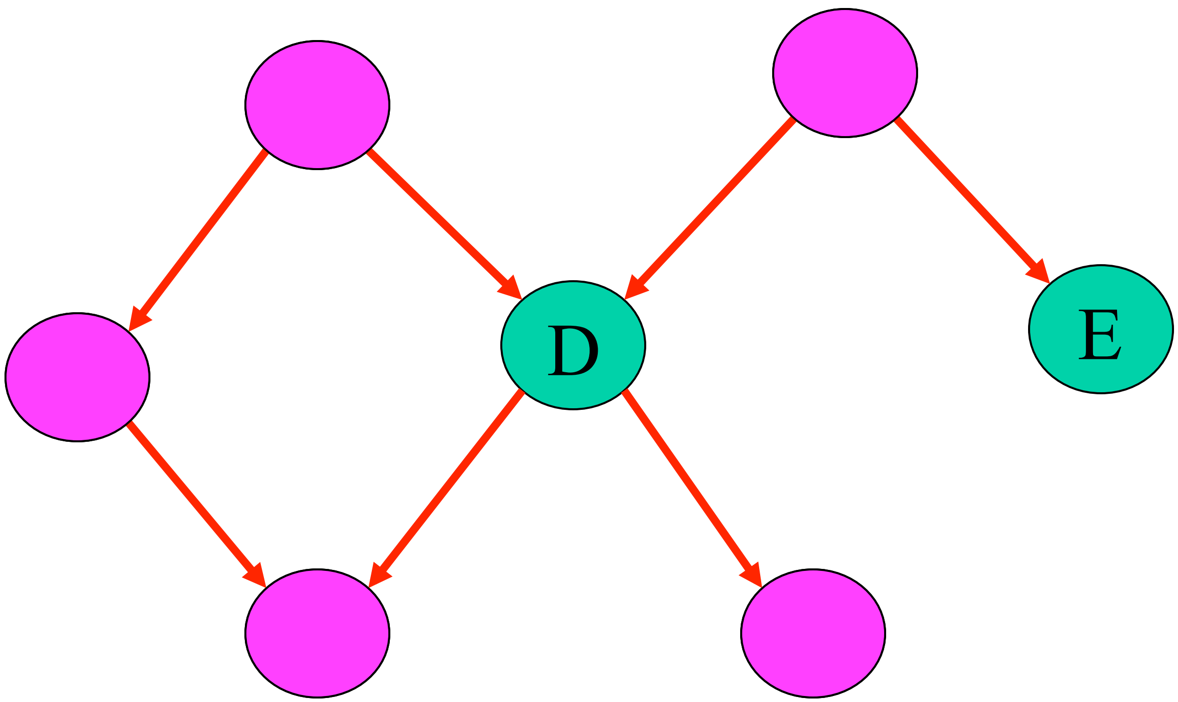


Pa(A) = 0
Pa(B) = 0
Pa(C) = A
Pa(D) = A, B
Pa(E) = B
Pa(F) = C, D
Pa(G) = D

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



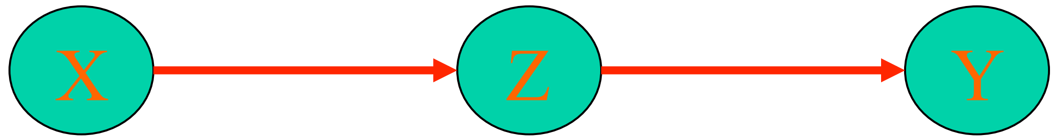
CM (D) ?

Axiomas de Independencia

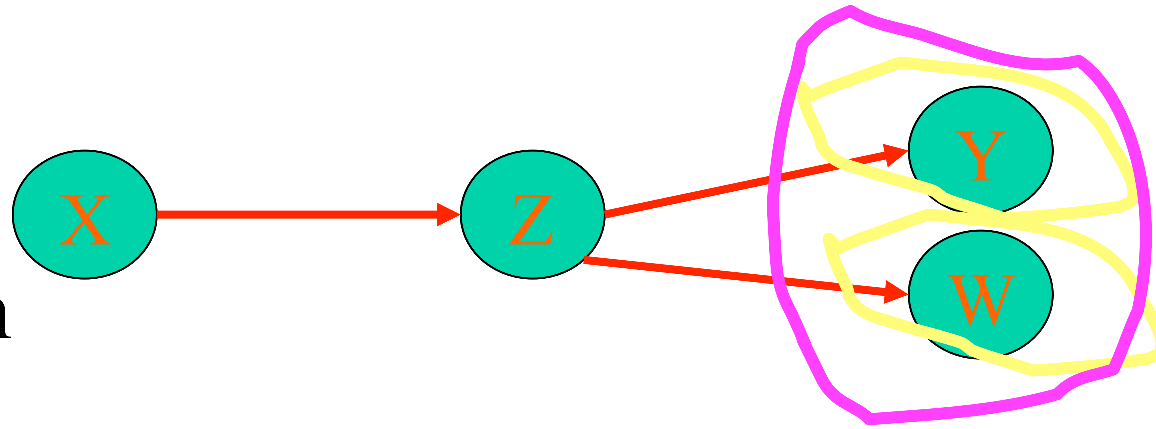
- A partir de ciertas relaciones de independencia se pueden derivar otras, sin necesidad de evaluar las probabilidades
- Para esto se pueden utilizar ciertas reglas. Las reglas básicas se conocen como axiomas de independencia.

Axiomas de Independencia

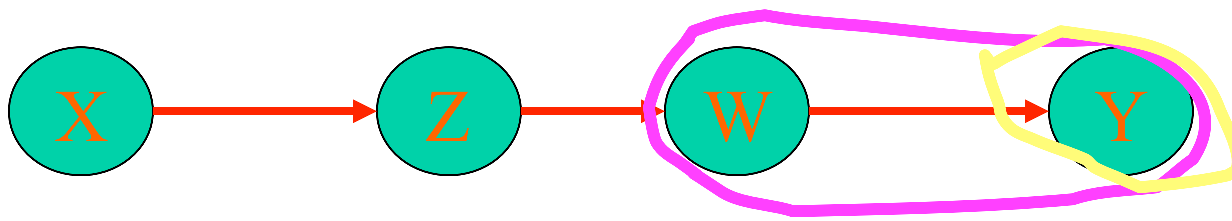
- Simetría



- Descomposición

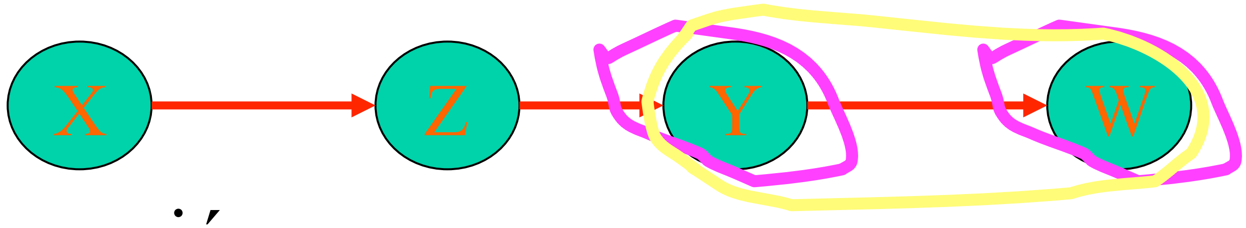


- Unión débil

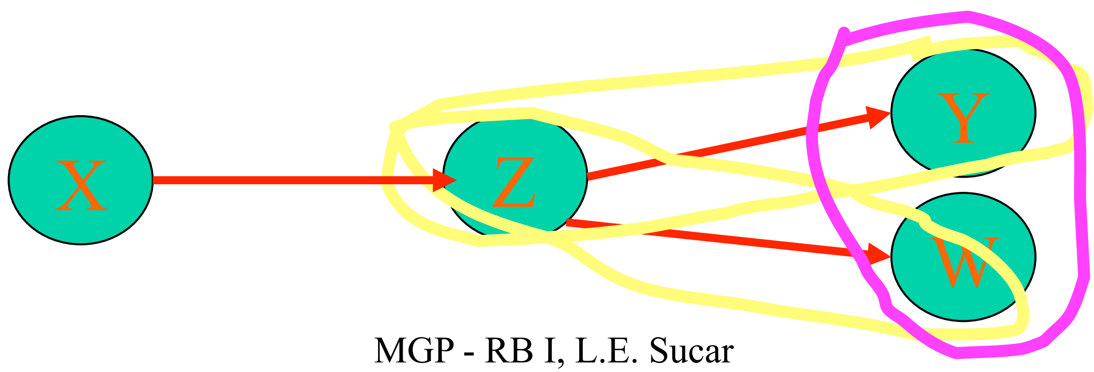


Axiomas de Independencia

- Contracción

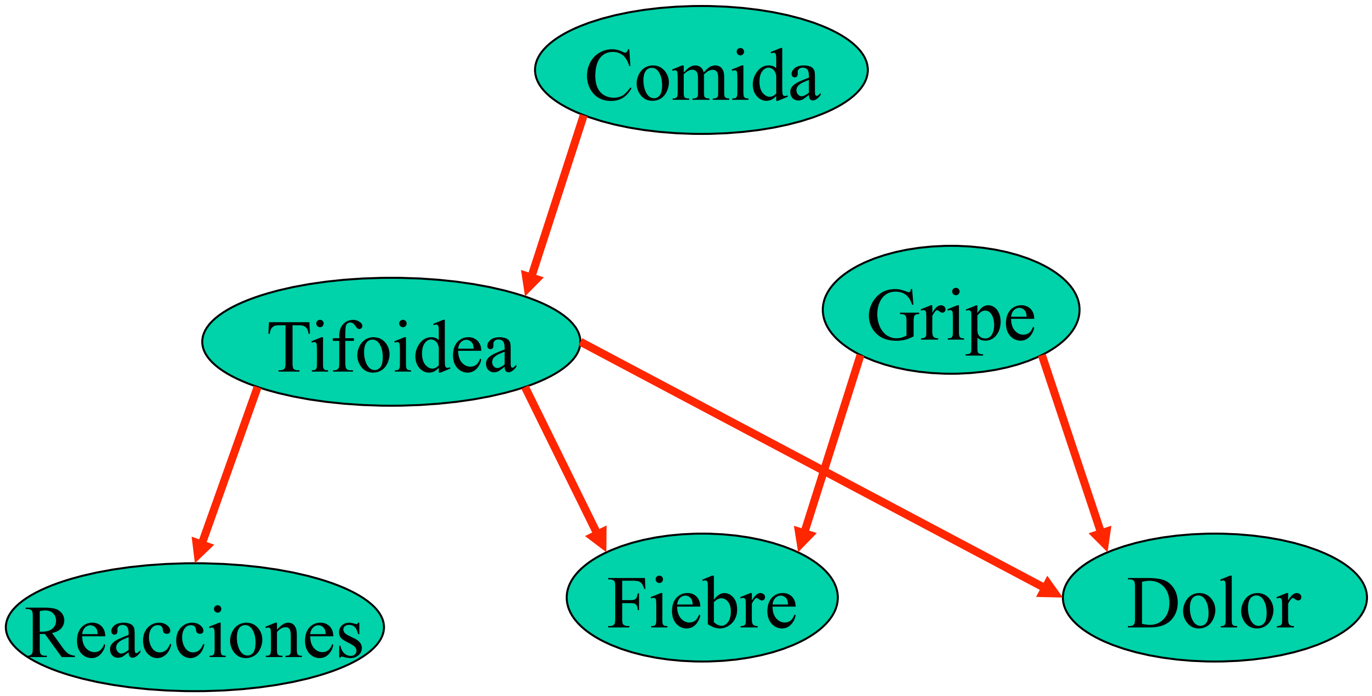


- Intersección

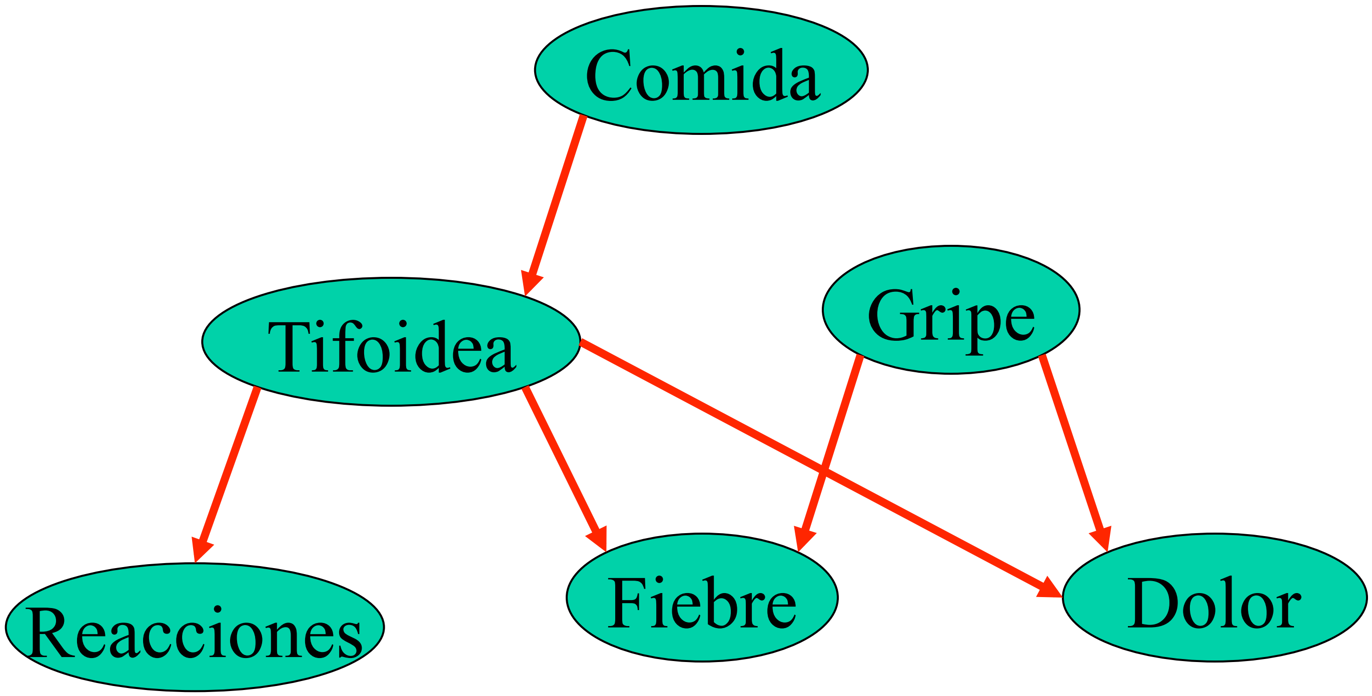


Parámetros

- 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


Comida
Tifoidea
Gripe
Reacciones
Fiebre
Dolor

Ins	Sal
0.2	0.8

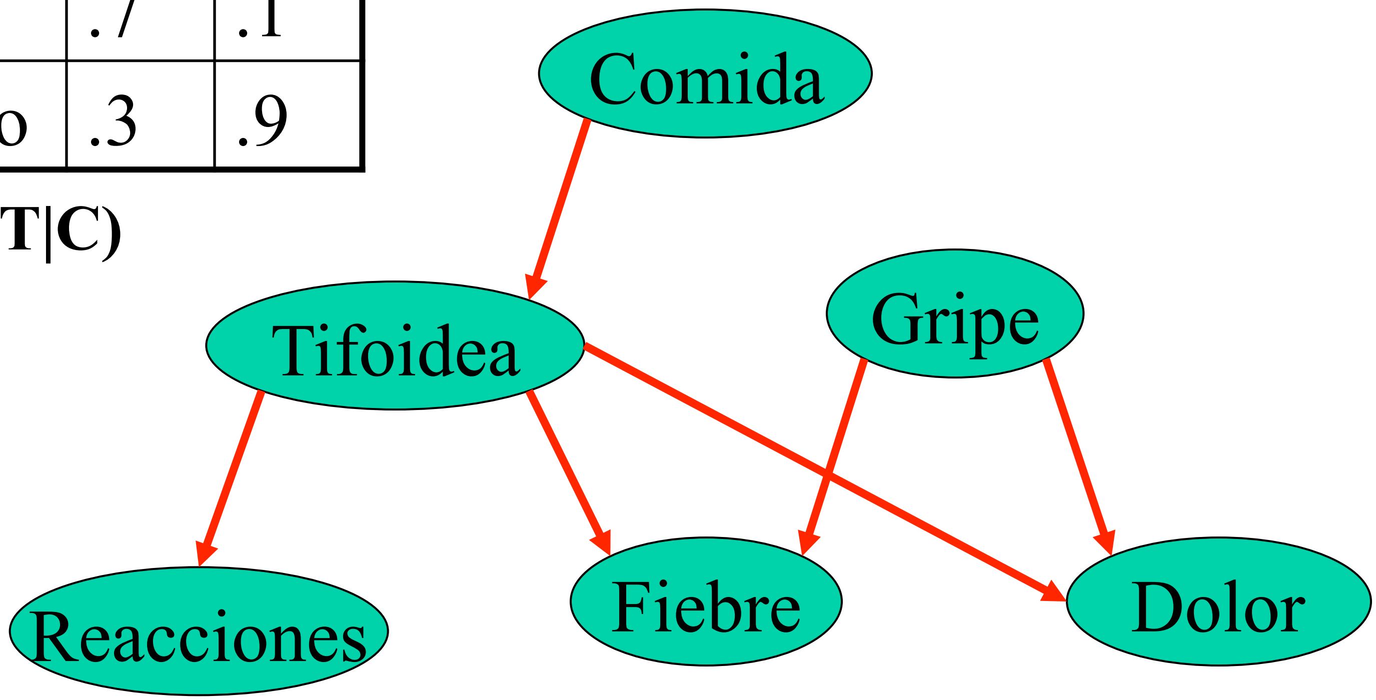


P(C)

Ins	Sal
0.2	0.8

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

P(T|C)

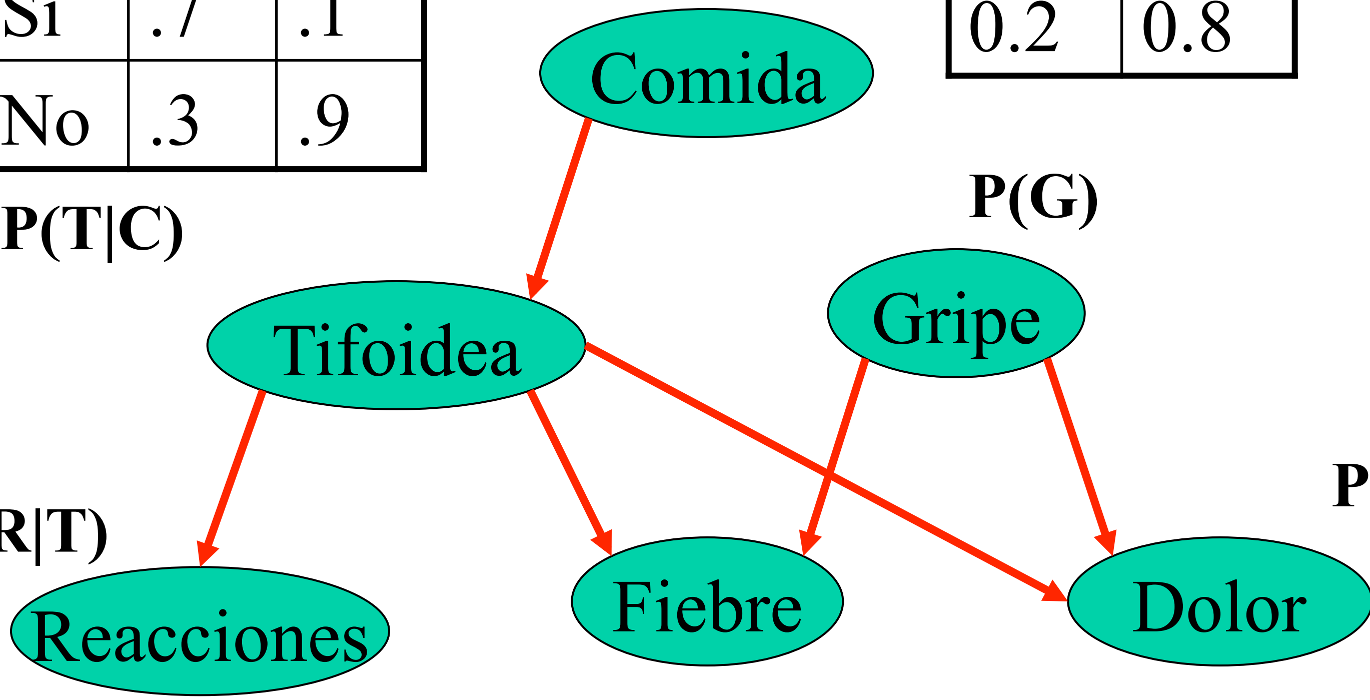


P(C)

Ins	Sal
0.2	0.8

P(T|C)

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



P(G)

P(R|T)

P(D|T,G)

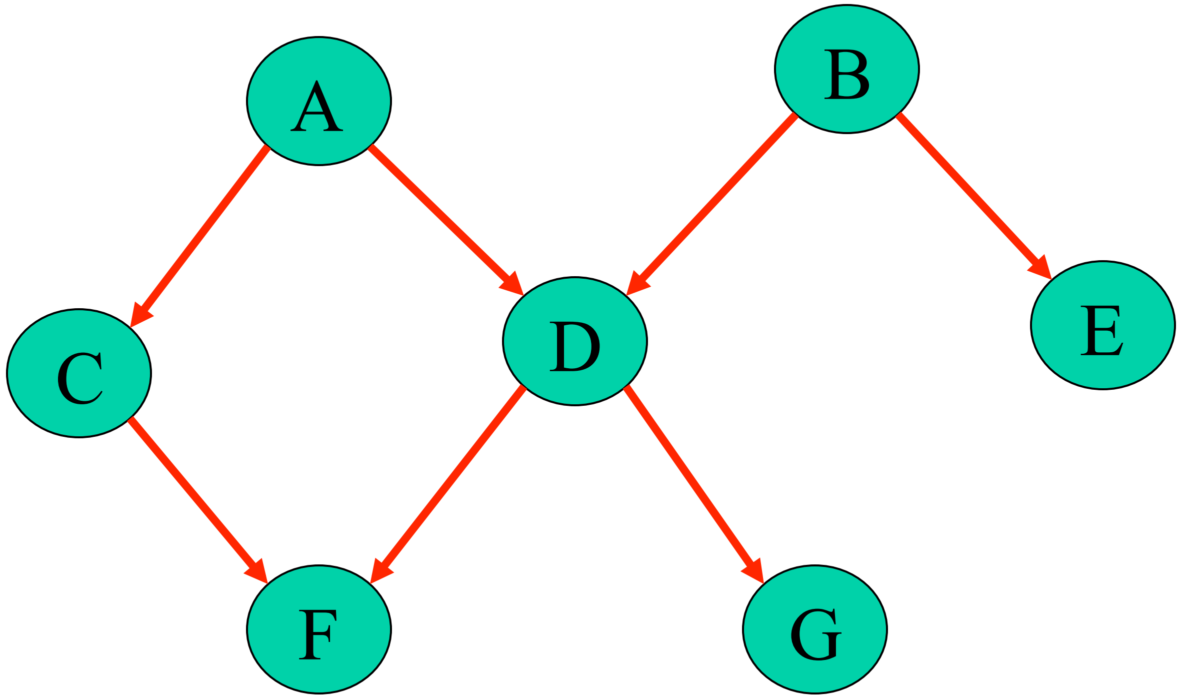
P(F|T,G)

	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)

= 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)

= 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:

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

Ejemplo de una red baysiana

HUGIN

Modelos canónicos

- El tamaño de la tabla de probabilidad condicional crece exponencialmente con el número de padres de un nodo, por lo que puede crecer demasiado
- Una forma de reducir este problema es utilizando ciertos modelos para representar las tablas sin requerir especificar todas las probabilidades – *modelos canónicos*

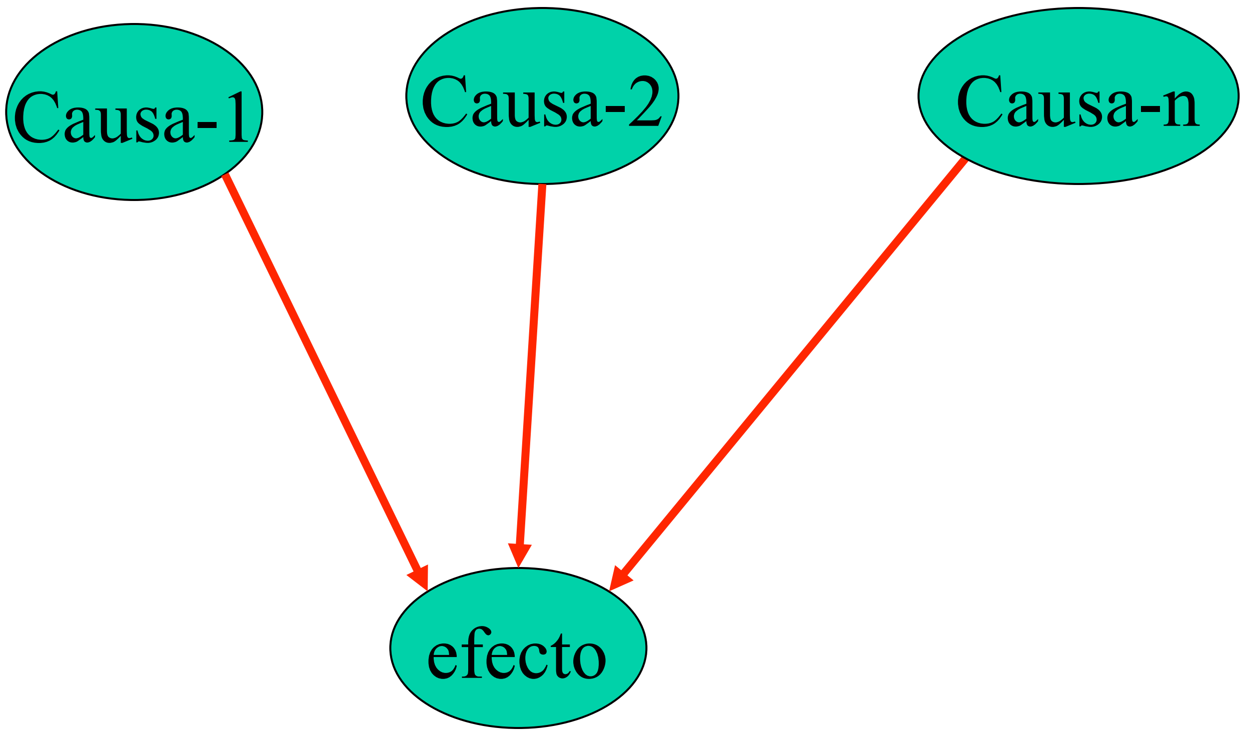
Modelos canónicos

- Tipos de modelos:
  - Modelo de interacción disjuntiva (*Noisy OR*)
  - Modelo de interacción conjuntiva (*Noisy AND*)
  - Compuerta Max (*Noisy Max gate*)
  - Compuerta Min (*Noisy Min gate*)

Noisy OR

- Se aplica cuando varias "causas" pueden ocacionar un "efecto" c/u por si sola, y la probabilidad del efecto no disminuye si se presentan varias causas
- Se considera que todas las variables son binarias
- Por ejemplo, este modelo se puede aplicar cuando varias enfermedades pueden producir el mismo síntoma

Noisy OR



Noisy OR

- Propiedades:
  - Responsabilidad: el efecto es falso si todas sus posibles causas son falsas
  - Independencia de excepciones: si un efecto es la manifestación de varias causas, los mecanismos que inhiben la ocurrencia del efecto bajo una causa, son independientes de los que lo inhiben bajo otras causas

Noisy OR

- Probabilidades:
  - Probabilidad de que el efecto sea inhibido por la causa i:

$$qi = P(\neg E \mid Ci)$$

  - En base a esto se puede calcular la matriz de probabilidad condicional como:

    $P(E \mid C1, ...Cn) =$

$$\prod_{Ci=1} qi, \qquad E=0$$

$$1 - \prod_{Ci=1} qi, \qquad E=1$$

Noisy OR

- Ejemplo: 3 causas, q1=q2=q3=0.1

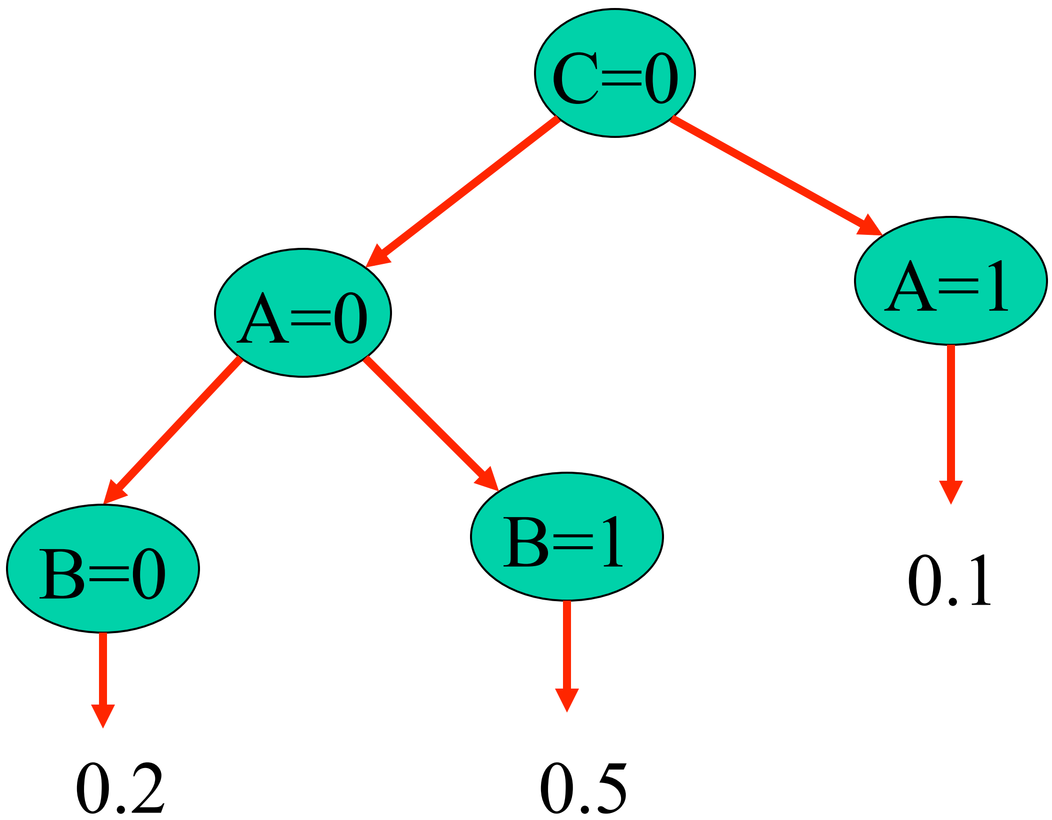
P(E | C1, C2, C3)

C1	0	0	0	0	1	1	1	1
C2	0	0	1	1	0	0	1	1
C3	0	1	0	1	0	1	0	1
E=0	1	.1	.1	.01	.1	.01	.01	.001
E=1	0	.9	.9	.99	.9	.99	.99	.999

Otras representaciones

- Otras formas compactas de representar las tablas de probabilidad condicional:
  - Árboles de decisión
  - Diagramas de decisión
  - Redes neuronales

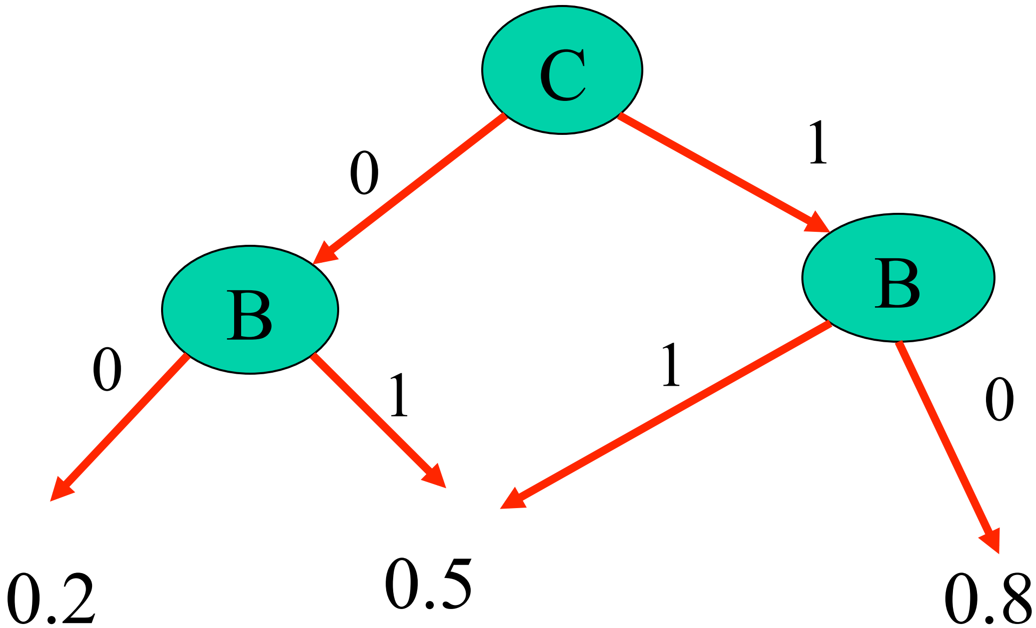
CPT – árbol de decisión



P(C|A,B)

00	01	10	11
0.2	0.5	0.1	0.1
0.8	0.5	0.9	0.9

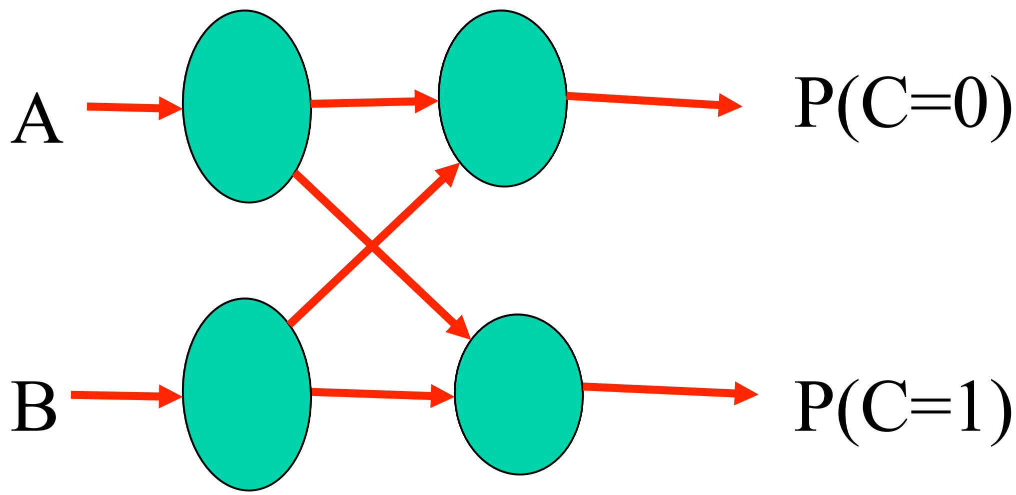
CPT – diagrama de decisión



P(C|A,B)

00	01	10	11
0.2	0.5	0.2	0.5
0.8	0.5	0.8	0.5

CPT – red neuronal



P(C|A,B)

00	01	10	11
0.2	0.5	0.1	0.1
0.8	0.5	0.9	0.9

Referencias

- Pearl 88 – Cap. 3
- Neapolitan 90 – Cap. 5
- Koller & Friedman - Cap. 3
- Sucar, Morales, Hoey - Cap. 2
- Sucar, L.E., “Introducción a redes bayesianas”, en Sierra (ed.), Aprendizaje Automático, Pearson, 2008
- F. J. Díez y M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01. UNED, Madrid, 2006

Actividades

- Leer capítulo redes bayesianas (en la página)
- Hacer ejercicios de representación de redes bayesianas