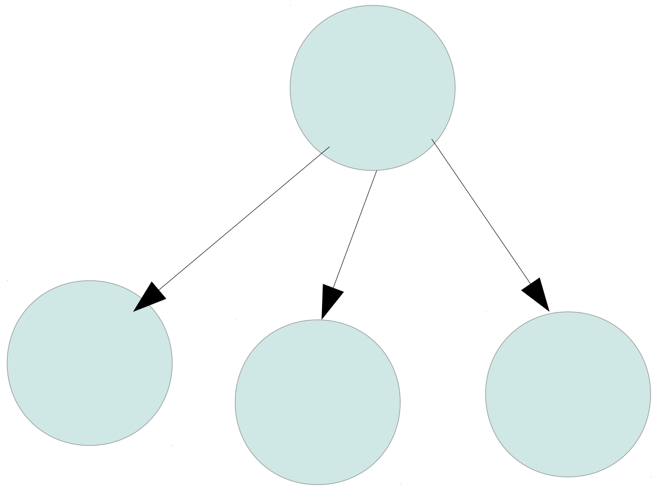
Razonamiento probabilístico y redes bayesianas



Semántica de las RB
Inferencia exacta

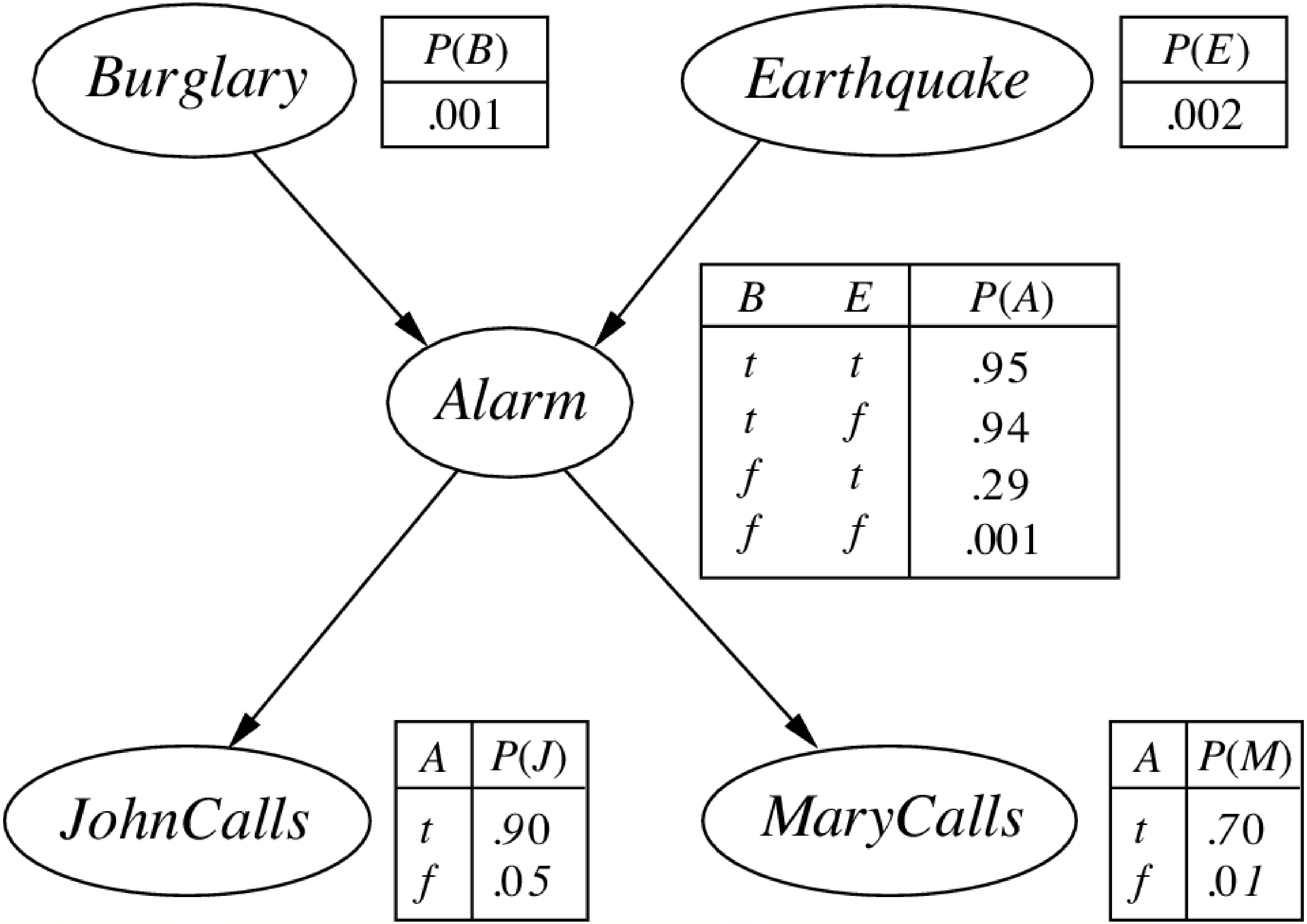
Blanca A. Vargas Govea
vargasgovea@itesm.mx

Octubre 16, 2012

Objetivos

- Conocer la semántica de las redes bayesianas.
- Conocer la inferencia exacta en redes bayesianas.

Ejemplo



$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

Semántica de las RB

Hay dos formas en las que se pueden entender

1. Pueden verse como una representación de la distribución de probabilidad conjunta.
2. También como una codificación de una colección de enunciados de independencia condicional.

Aunque son equivalentes, la primera nos sirve para entender cómo construir las redes mientras que la segunda nos ayuda a diseñar métodos de inferencia.

Representando la distribución conjunta completa

- Una RB proporciona una descripción completa del dominio.
- Cada entrada puede calcularse a partir de la información en la red.
- Una entrada es la probabilidad de una conjunción de asignación de valores a cada variable $P(X_1 = x_1 \cap \ldots \cap X_n = x_n)$

que se abrevia como: $P(x_1, \ldots, x_n)$

Representando la distribución conjunta completa

- Cada entrada en la distribución conjunta se representa por el producto de los elementos apropiados de las tablas (CPT).

$$P(x_{1}, \ldots, x_{n}) = \prod_{i=1}^{n} P(x_i | padres(X_i))$$

Representando la distribución conjunta completa

- Podemos calcular la probabilidad de que la alarma haya sonado pero no hay ni robo ni temblor y que tanto Mary como John llamen:

$$P(j \cap m \cap a \cap \sim b \cap \sim e) =$$
$$P(j|a)P(m|a)P(a|\sim b \cap \sim e)P(\sim b)P(\sim e) =$$
$$0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$

Construcción de RBs

La ecuación

$$P(x_{1}, \ldots, x_{n}) = \prod_{i=1}^{n} P(x_i | padres(X_i))$$

implica ciertas relaciones de independencia condicional que pueden ser usadas para construir la RB.

Construcción de RBs

Se reescribe la probabilidad conjunta en términos de una probabilidad condicional

$P(a \cap b) = P(a|b)\, P(b)$

$P(x_1,\ldots,x_n) =$

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

$$\prod_{i=1}^{n} P(x_i|x_{i-1},\ldots,x_1)$$

Regla de la cadena

Regla de la cadena: ejemplo

$$P(x_1,x_2,x_3,x_4) =$$

$$P(x_1)\,P(x_2|x_1)\,P(x_3|x_2,x_1)\,P(x_4|x_3,x_2,x_1)$$

$$\prod_{i=1}^{n} P(x_i|x_{i-1},\ldots,x_1)$$

Una RB es una representación correcta del dominio sólo si cada nodo es condicionalmente independiente de sus predecesores dados sus padres.

Construcción de RBs

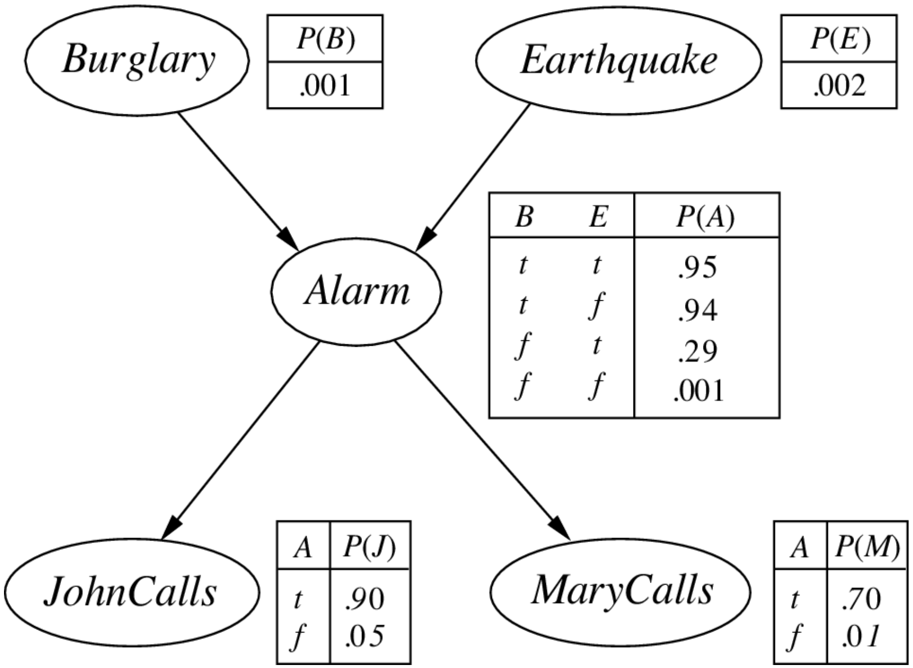
1. Nodos: determinar el no. de variables requeridas para modelar el dominio. La red será más compacta si las variables se ordenan de modo que las causas precedan a los efectos.
2. Links: estructura acíclica. Cualquier variable no-descendiente del nodo X es independiente condicionalmente de X dados sus padres.
3. Agregar tabla de probabilidad para cada nodo.

Construcción de RBs

- Causa → Efecto. Se modelan menos ligas.
- Al revés se tienen que especificar dependencias condicionales.

P(MaryLlama|
JohnLlama,
Alarma,Temblor,
Robo) =

P(MaryLlama|Alarma)



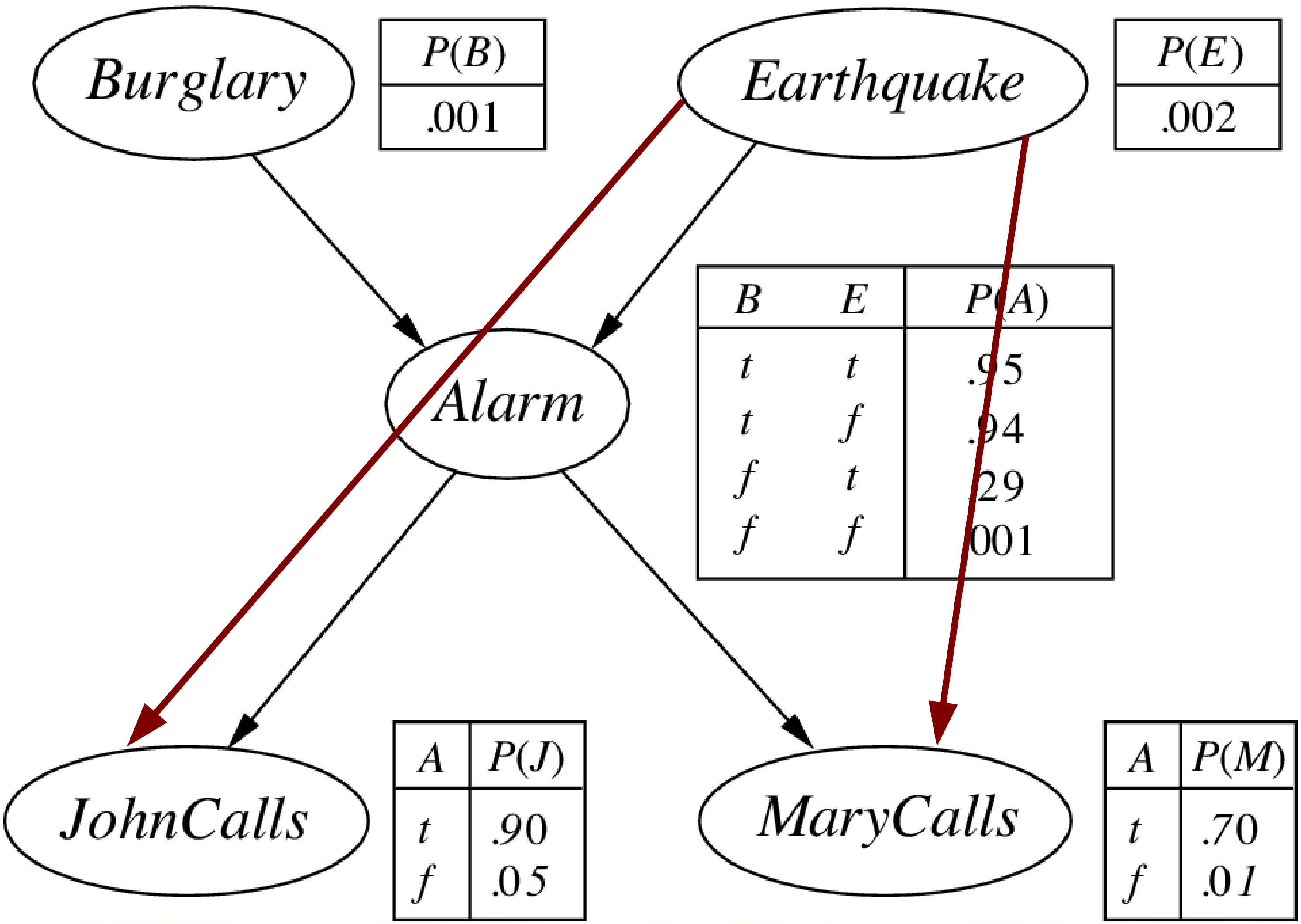
Compresión y orden de nodos

- Una RB es a menudo más compacta que la distribución conjunta completa.
- Sistema localmente estructurado. Cada componente interactúa con un número limitado de componentes.
- Estructura local: crecimiento lineal.

Compresión y orden de nodos

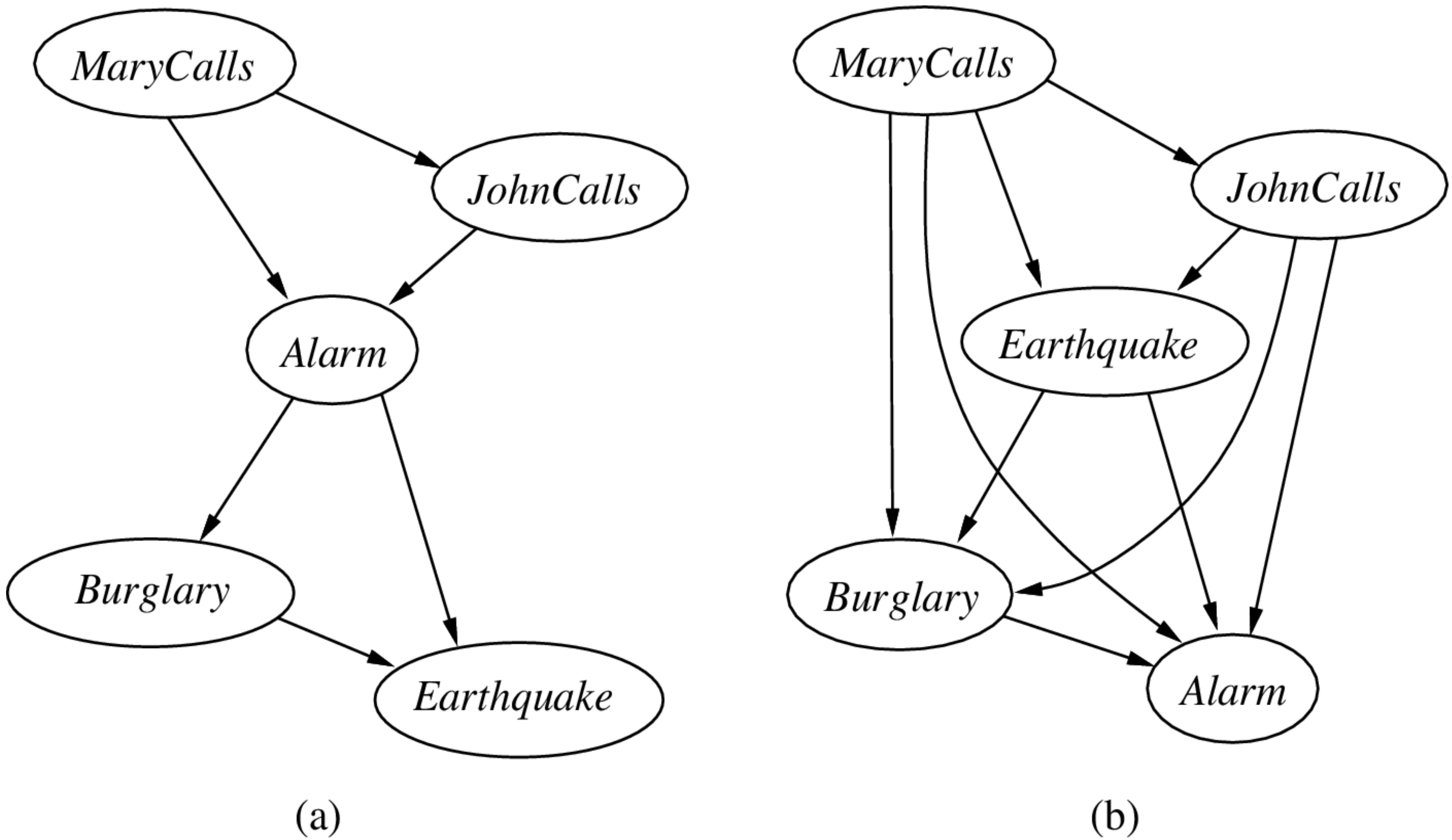
- Rbs: cada variable aleatoria es directamente influenciada por k nodos.
- n variables booleanas.
- 2^k números: información necesaria para especificar la tabla; $n2^k$: red completa.
- Distribución conjunta: 2^n
- Para $n=30$ y $k=5$, RB requiere $(30)2^5=960$ números y la distribución conjunta 2^{30} $=1,073,741,824$.

Si Mary y John saben que tembló no llamarán. ¿Agregar la liga?



Orden de nodos

- Determina la compresión de la red.
- Agregar las causas primero, luego las variables que influencian hasta llegar a las hojas.
- Las hojas ya no tienen influencia causar en otras variables.



Más ligas que la red original. Las relaciones no son claras.

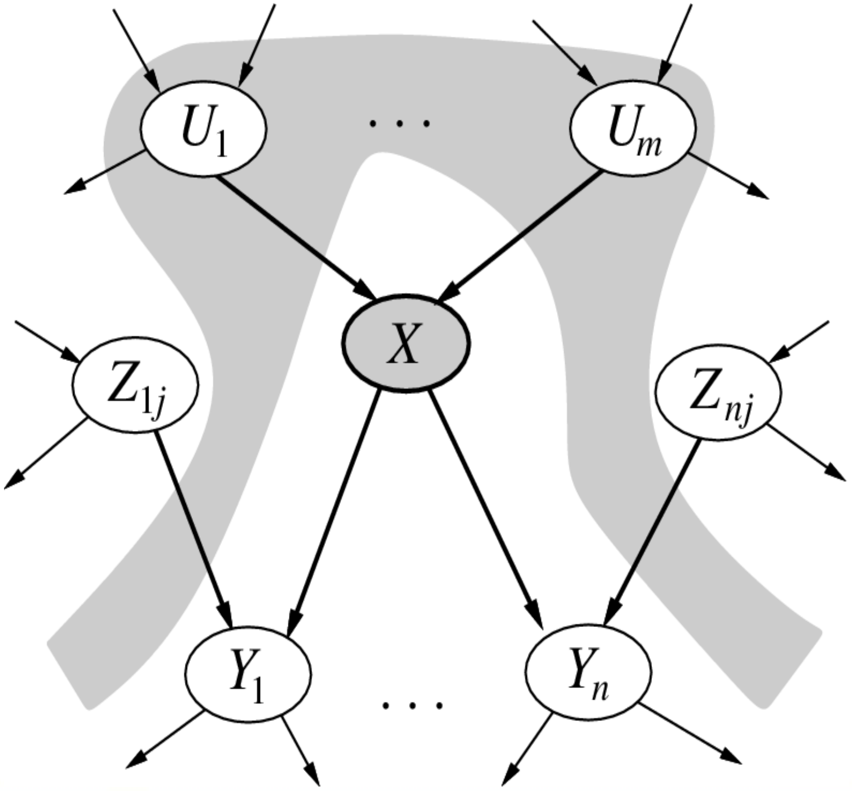
Relaciones de independencia condicional

En vez de obtener las relaciones de independencia condicional a partir de la distribución conjunta se parte de la topología.

Semántica de la topología

1. Cada variable es condicionalmente independiente de sus no-descendientes dados sus padres.

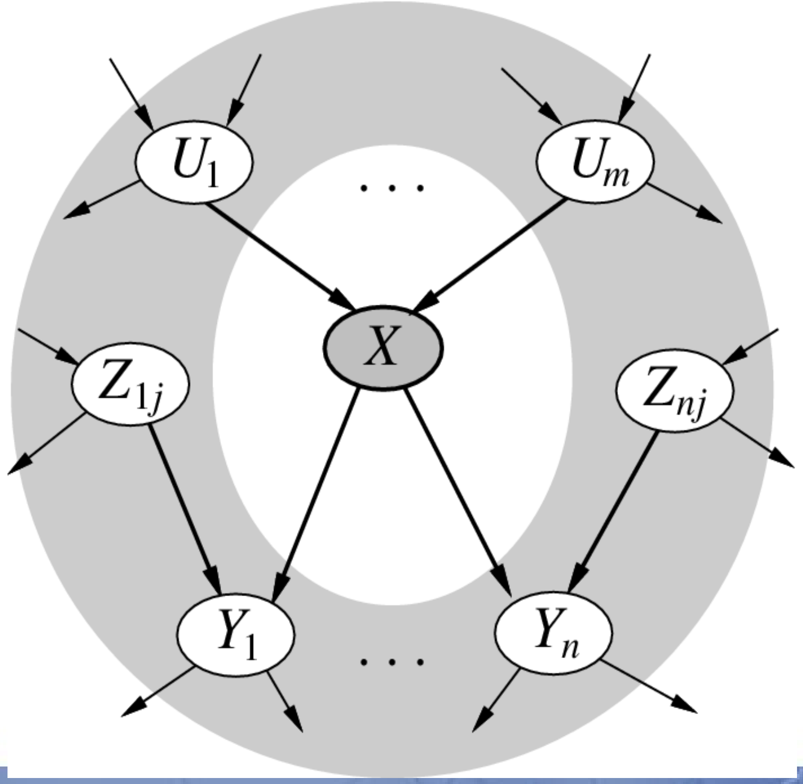
JohnLlama es independiente de Robo, Temblor y MaryLlama dado el estado de la Alarma.



Semántica de la topología

2. Un nodo es condicionalmente independiente del resto de nodos de la red dados sus padres, hijos y padres de sus hijos.

Robo es independiente de JohnLlama y MaryLlama dado Alarma y Temblor.



Inferencia exacta:inferencia por enumeración

- La tarea básica de un sistema de información probabilista es calcular la distribución de probabilidad posterior para un conjunto de variables dado un evento.
- La distribución conjunta puede escribirse en términos de productos de probabilidades condicionales.

Inferencia exacta:inferencia por enumeración

Un query puede contestarse utilizando una red Bayesiana calculando la suma de productos de las probabilidades condicionales de la red.

Calcular la probabilida de que un robo haya ocurrido:

P(Robo|JohnLlama=true,MaryLlama=true)

Inferencia exacta:inferencia por enumeración

P(Robo|JohnLlama=true,MaryLlama=true)

de la ec.

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X,\mathbf{e}) = \alpha\sum_{\mathbf{y}}\mathbf{P}(X,\mathbf{e},\mathbf{y})$$

suma sobre las posibles combinaciones en las variables no observadas

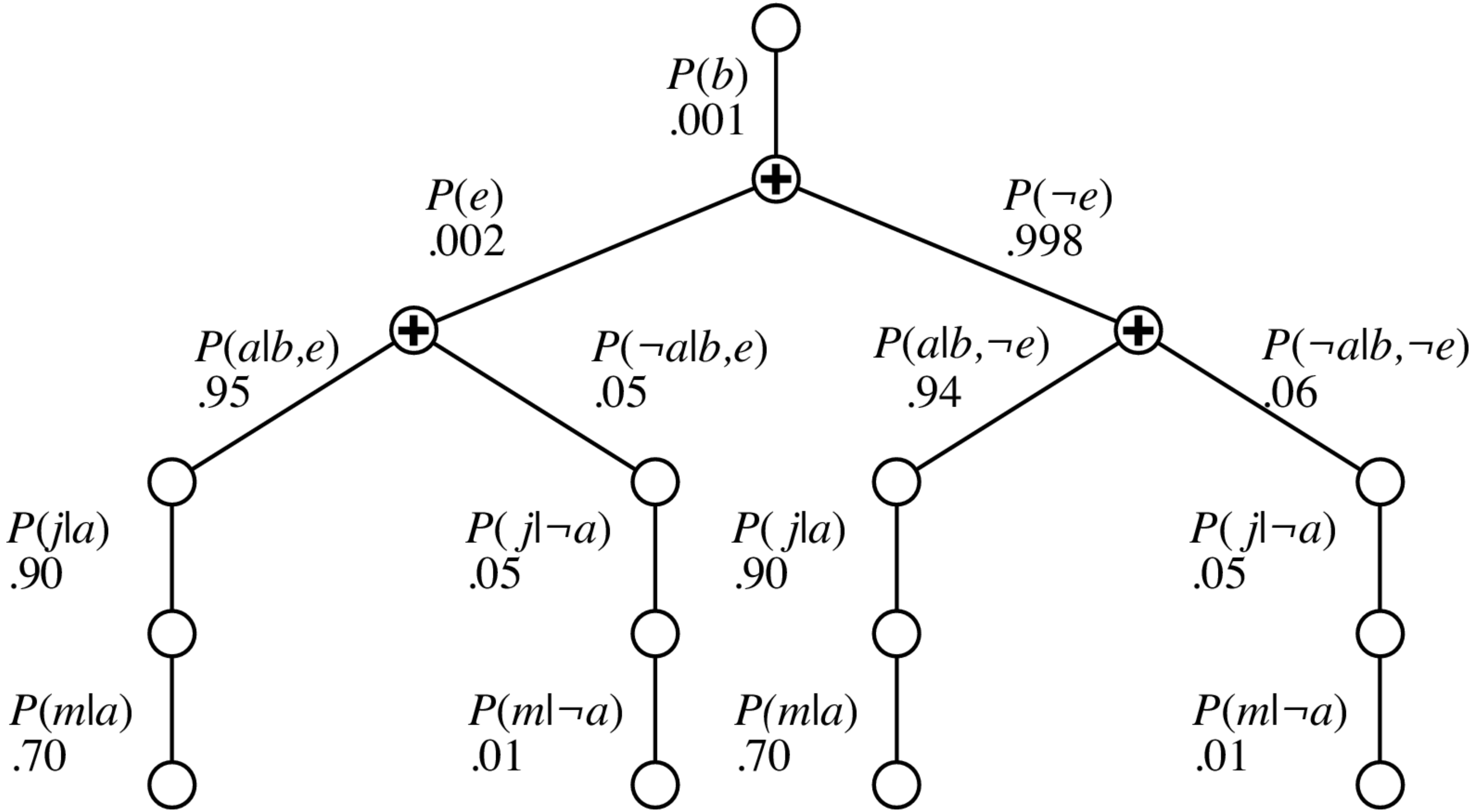
$$\mathbf{P}(B|j,m) = \alpha\mathbf{P}(B,j,m) = \alpha\sum_{e}\sum_{a}\mathbf{P}(B,j,m,e,a,)$$

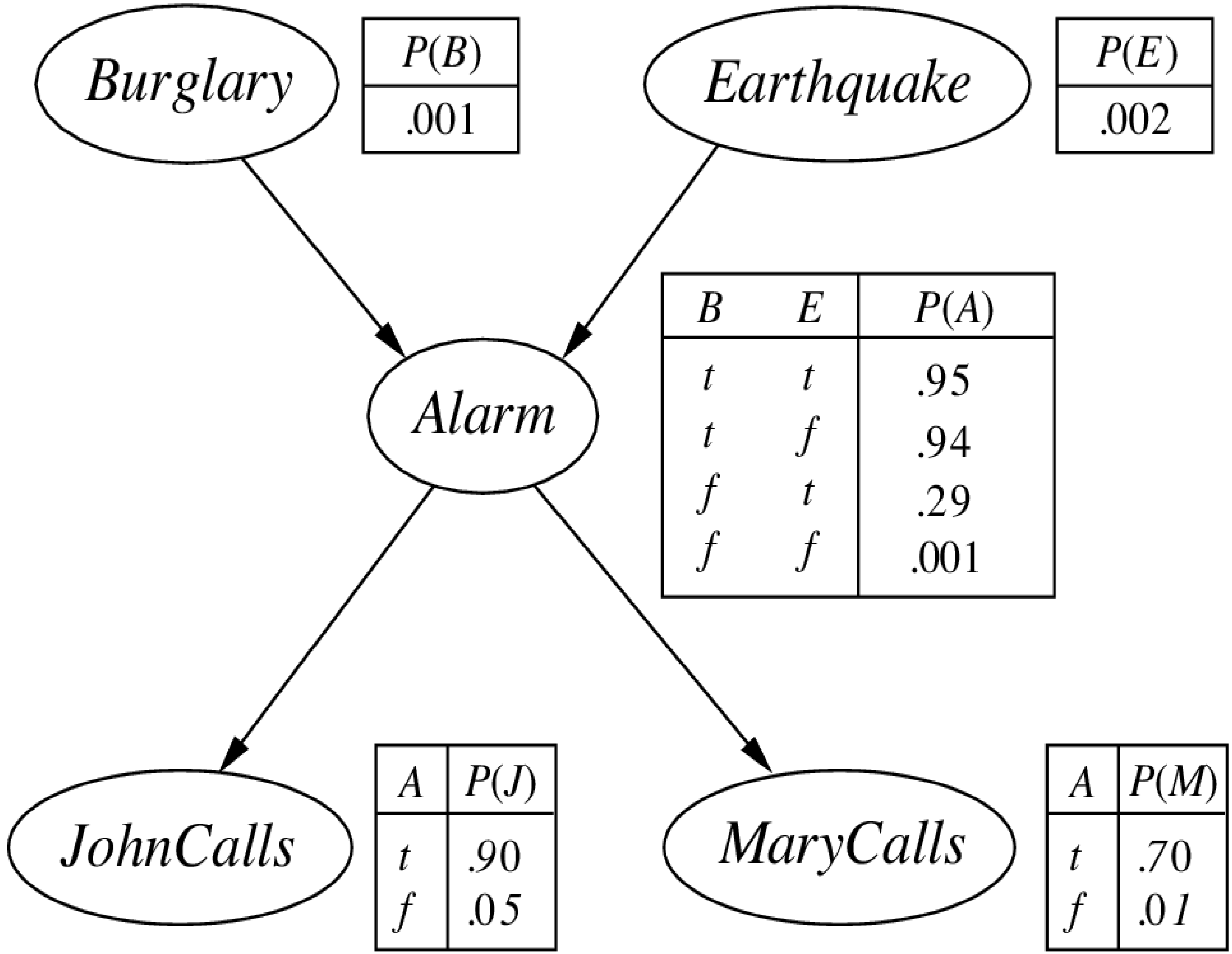
la semántica de RB permite expresar en términos de independencia condicional

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

$$P(b|j,m) = \alpha\sum_{e}\sum_{a}P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$

$$P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) .$$





$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

Inferencia exacta:inferencia por enumeración

$p(b|j,m)=$
$p(b)p(e)p(a|b,e)p(j|a)p(m|a)+$
$p(b)p(e)p(\sim a|b,e)p(j|\sim a)p(m|\sim a)+$
$p(b)p(\sim e)p(a|b,\sim e)p(j|a)p(m|a)+$
$p(b)p(\sim e)p(\sim a|b,\sim e)p(j|\sim a)p(m|\sim a)$

De las tablas:

$(0.001)(0.002)(0.95)(0.90)(0.70)+$	$1.197\text{x}10^{-6}$
$(0.001)(0.002)(0.05)(0.05)(0.01)+$	$5\text{x}10^{-11}$
$(0.001)(0.998)(0.94)(0.90)(0.70)+$	$5.9\text{x}10^{-4}$
$0.001(0.998)(0.06)(0.05)(0.01)$	$\underline{2.994\text{x}10}^{-8}$
	0.0005922

Inferencia exacta:inferencia por enumeración

p(b|j,m)=0.0005922
p(~b|j,m)=0.0014919

Normalizando:
p(b|j,m) + p(~b|j,m) = 1
α(0.0005922) + α(0.0014919) = 1
α(0.0005922+0.0014919) = 1
despejando $\alpha = 4.79814 \times 10^{2}$

p(b|j,m)=0.284
p(~b|j,m)= 0.7158

La probabilidad de que se trate de un robo dado que los dos vecinos llamaron es de 28%

Ejercicio

Construye una red Bayesiana que represente el siguiente escenario:

- Considera 4 nodos binarios: nublado, aspersor, lloviendo y pasto-mojado.
- La mayor probabilidad de que el pasto esté mojado es cuando el aspersor está puesto y está lloviendo.
- La probabilidad es menor cuando solamente una de las evidencias se observa.
- La probabilidad de que esté nublado es 0.5.

Tarea

- Instala Hugin (versión demo) http://www.hugin.com/
- Construye la red del ejercicio.
- Compila y corrige en caso necesario.