

Raciocínio Probabilístico



Sumário

- Representação de conhecimento num domínio incerto
- Redes de Bayes (redes Bayesianas)
- Inferência em Redes Bayesianas

Distribuição de Probabilidade Conjunta (Joint Probability Distribution)

- A partir da distribuição de probabilidade conjunta podemos determinar qualquer probabilidade referente às variáveis aleatórias
 - $P(\text{dorDeDentes})$
 - $P(\text{dorDeDentes}|\text{carieNoDente})$

	dorDeDentes		¬dorDeDentes	
	Detecta	¬ Detecta	Detecta	¬ Detecta
carieNoDente	0.108	0.012	0.072	0.008
¬ carieNoDente	0.016	0.064	0.144	0.576

Distribuição de Probabilidade Conjunta (Joint Probability Distribution)

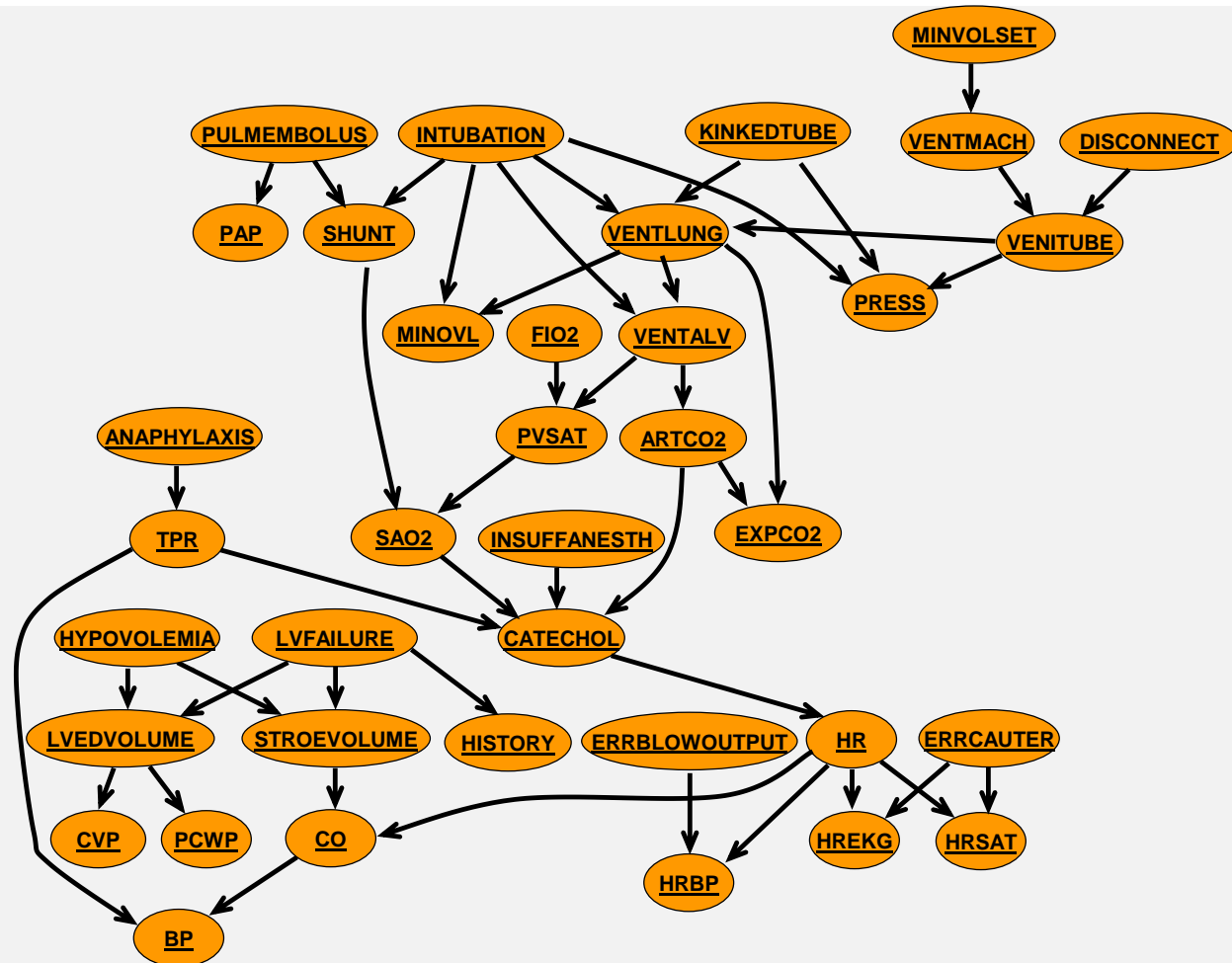
- No entanto a representação de uma Distribuição de Probabilidade Conjunta não é nada eficiente
 - seja n o número de variáveis aleatórias
 - e d o tamanho do domínio para cada variável aleatória
 - $O(d^n)$
 - $O(2^n)$ no caso de variáveis aleatórias booleanas

Redes Bayesianas

- Portanto precisamos de uma representação mais eficiente
 - mas que nos permita inferir qualquer tipo de probabilidades
- É para isso que servem as redes bayesianas

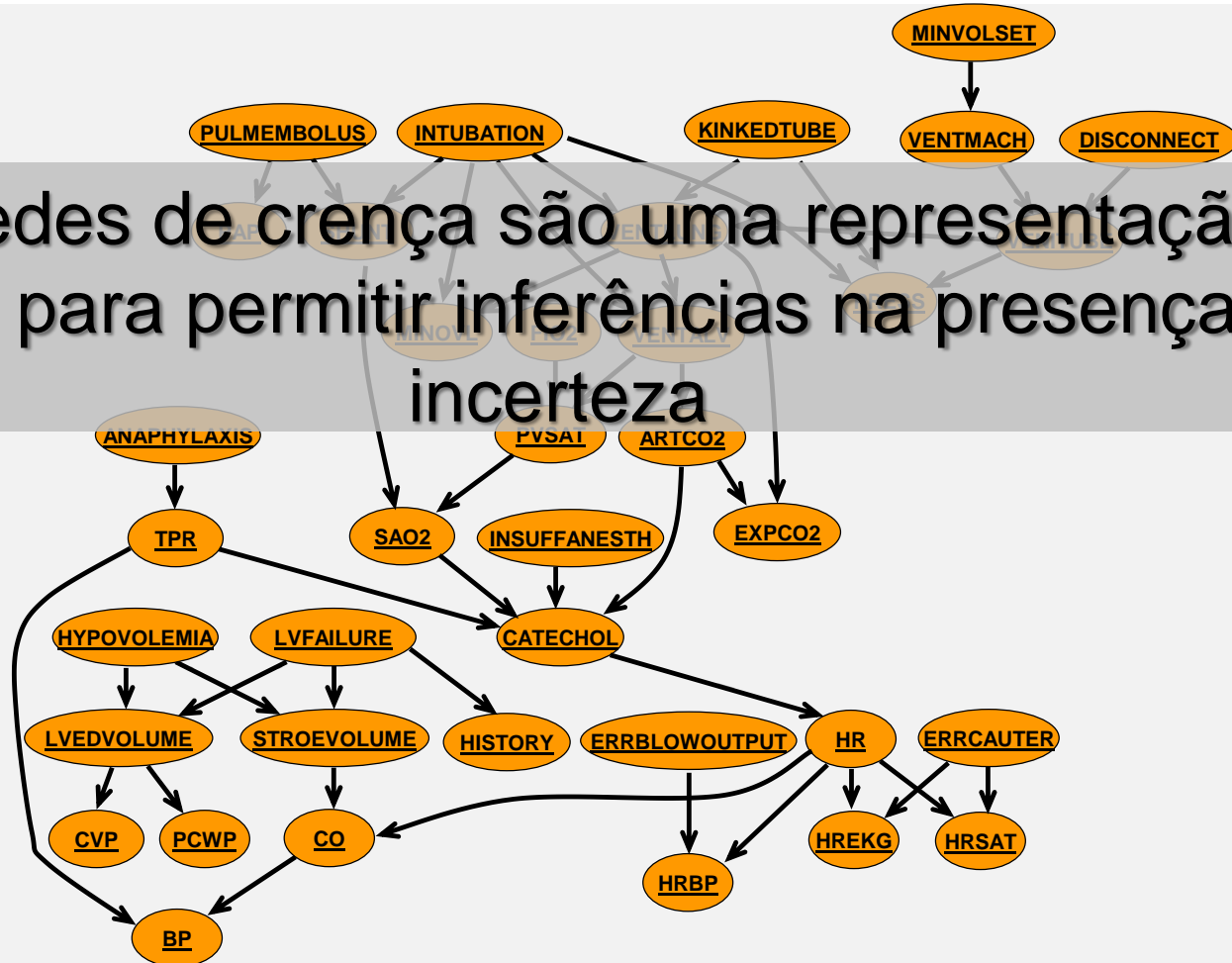
Redes Bayesianas (Redes de crenças) Bayesian Belief Networks

Redes Bayesianas

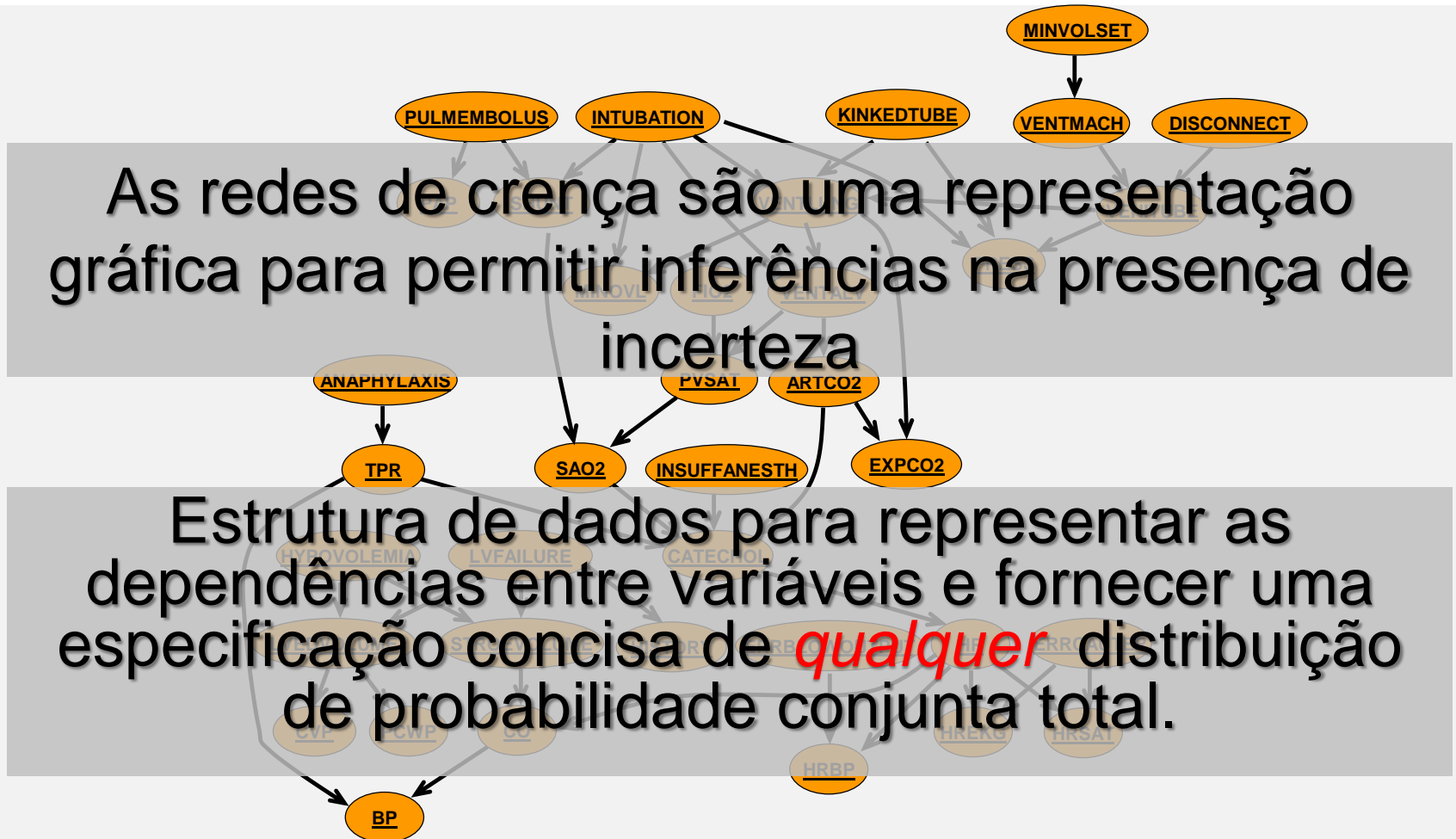


Redes Bayesianas

As redes de crença são uma representação gráfica para permitir inferências na presença de incerteza



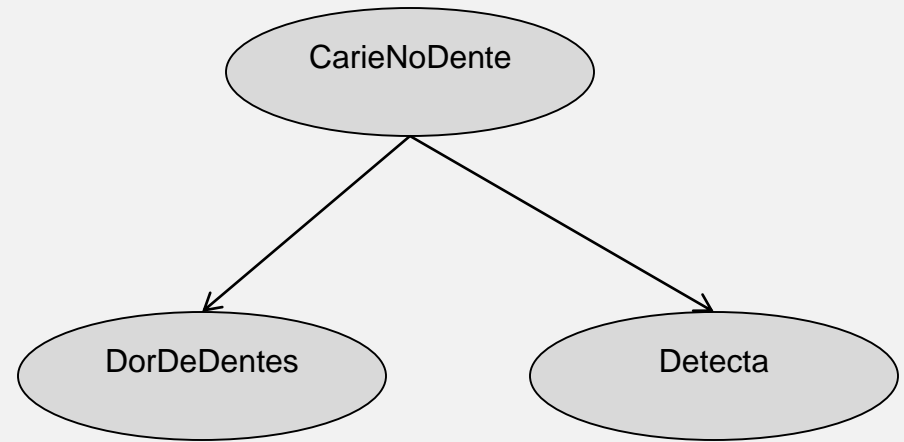
Redes Bayesianas



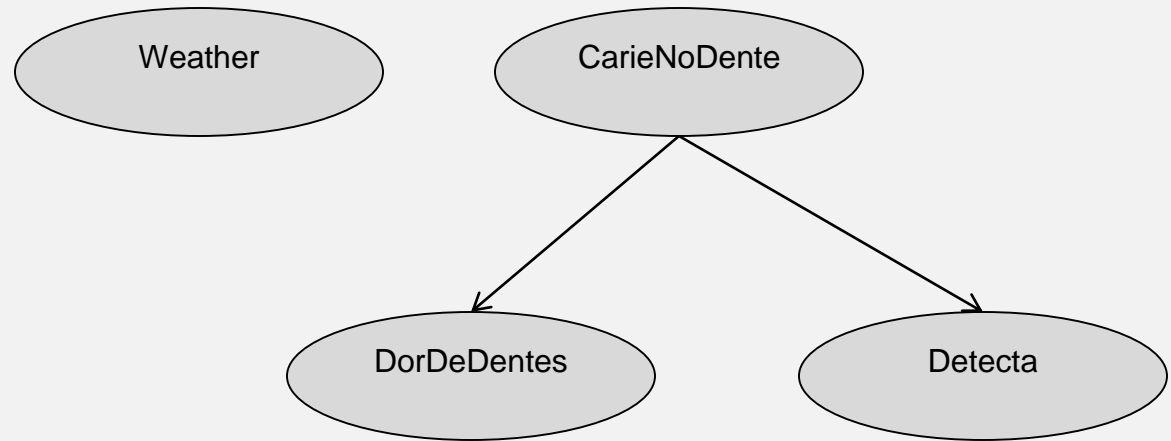
Redes Bayesianas

- Uma rede Baysiana é um grafo direccionado e acíclico em que cada nó é anotado com a informação sobre probabilidades.
- Sintaxe:
 - Cada nó corresponde a uma variável aleatória (que pode ser discreta ou contínua)
 - Contém um conjunto de ligações direccionadas (seta = "influência direta") de "pais" para filhos
 - cada nó tem uma distribuição condicional $P(X_i | \text{Pais}(X_i))$ que quantifica o efeito dos pais sobre o nó

Exemplo

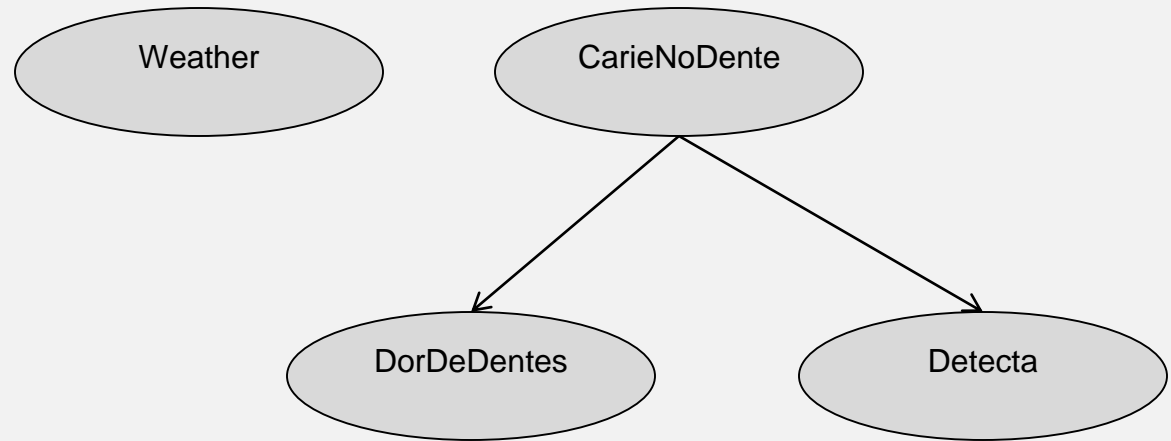


Exemplo



A topologia de uma rede representa relações de independência condicional

Exemplo



- “*Tempo*” é independente de outras variáveis
- *DorDeDentes* e *Detecta* são condicionalmente independentes dado *CarieNoDente*

Tabela de Probabilidade Condicional

- No caso mais simples, a distribuição condicional é representada como uma **tabela de probabilidade condicional** (TPC) dada uma distribuição sobre X_i para cada combinação de valores dos pais.

Exemplo

" I have two *neighbors* , *John and Mary*, who have promised to call when they hear the alarm. I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar? "

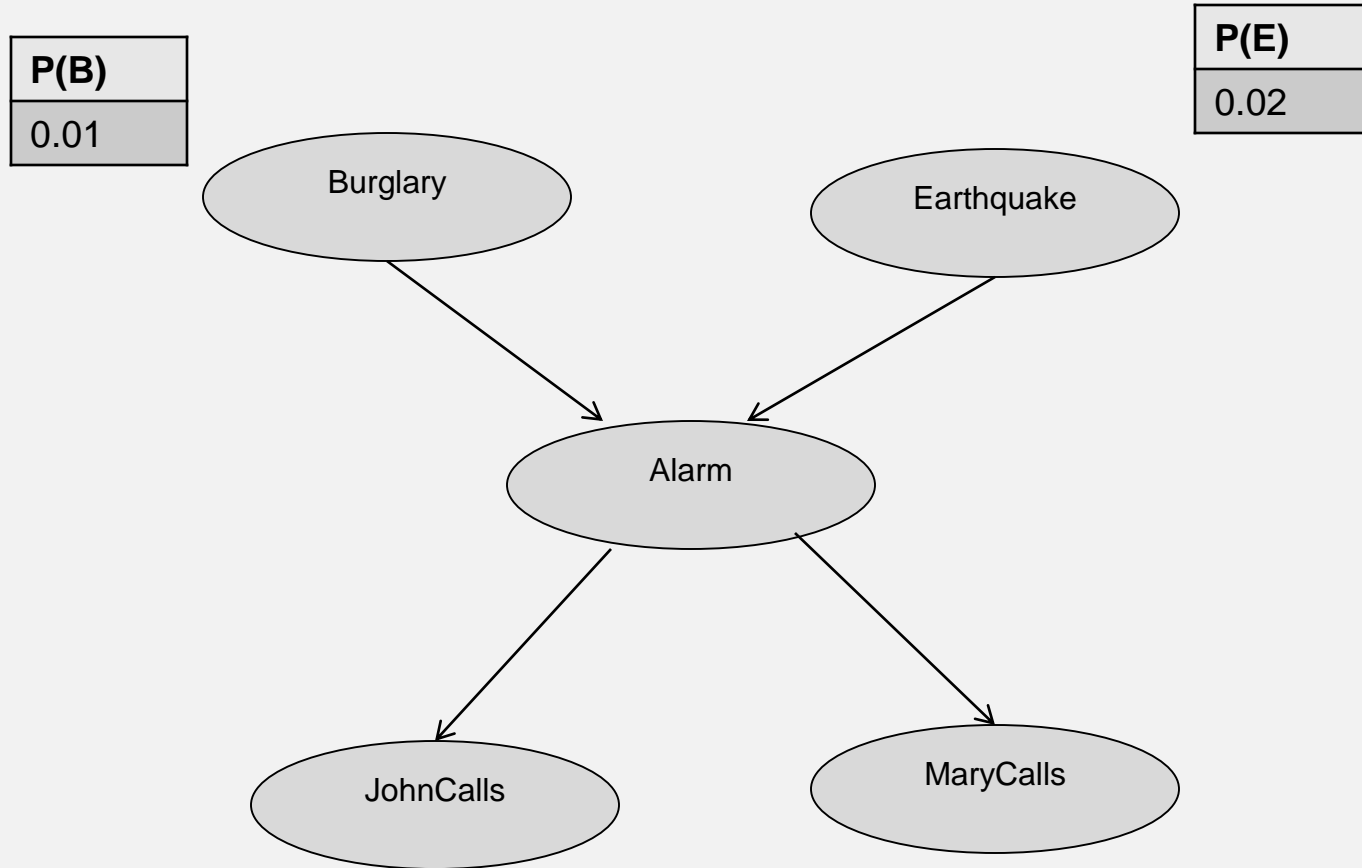


Exemplo

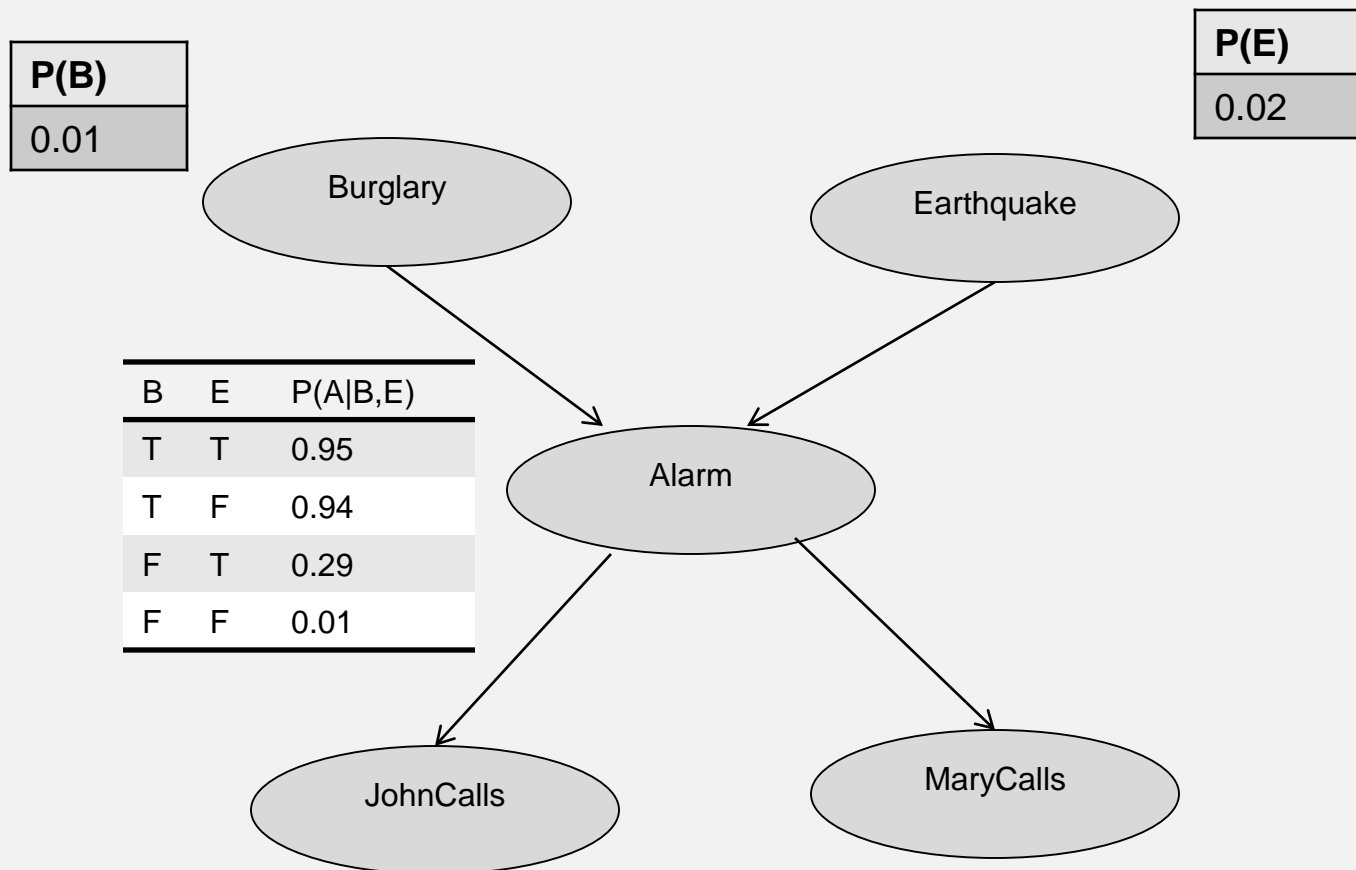
" I have two *neighbors* , *John and Mary*, who have promised to call when they hear the alarm. I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar? "

- Variáveis: *Burglary, Earthquake, Alarm, JohnCalls, MaryCalls*
- A topologia da rede reflete conhecimento "**causal**":
 - Um roubo (*burglary*) pode fazer disparar o alarme
 - Um terramoto (*earthquake*) pode ligar o alarme
 - O alarme faz Mary telefonar
 - O alarme faz John telefonar

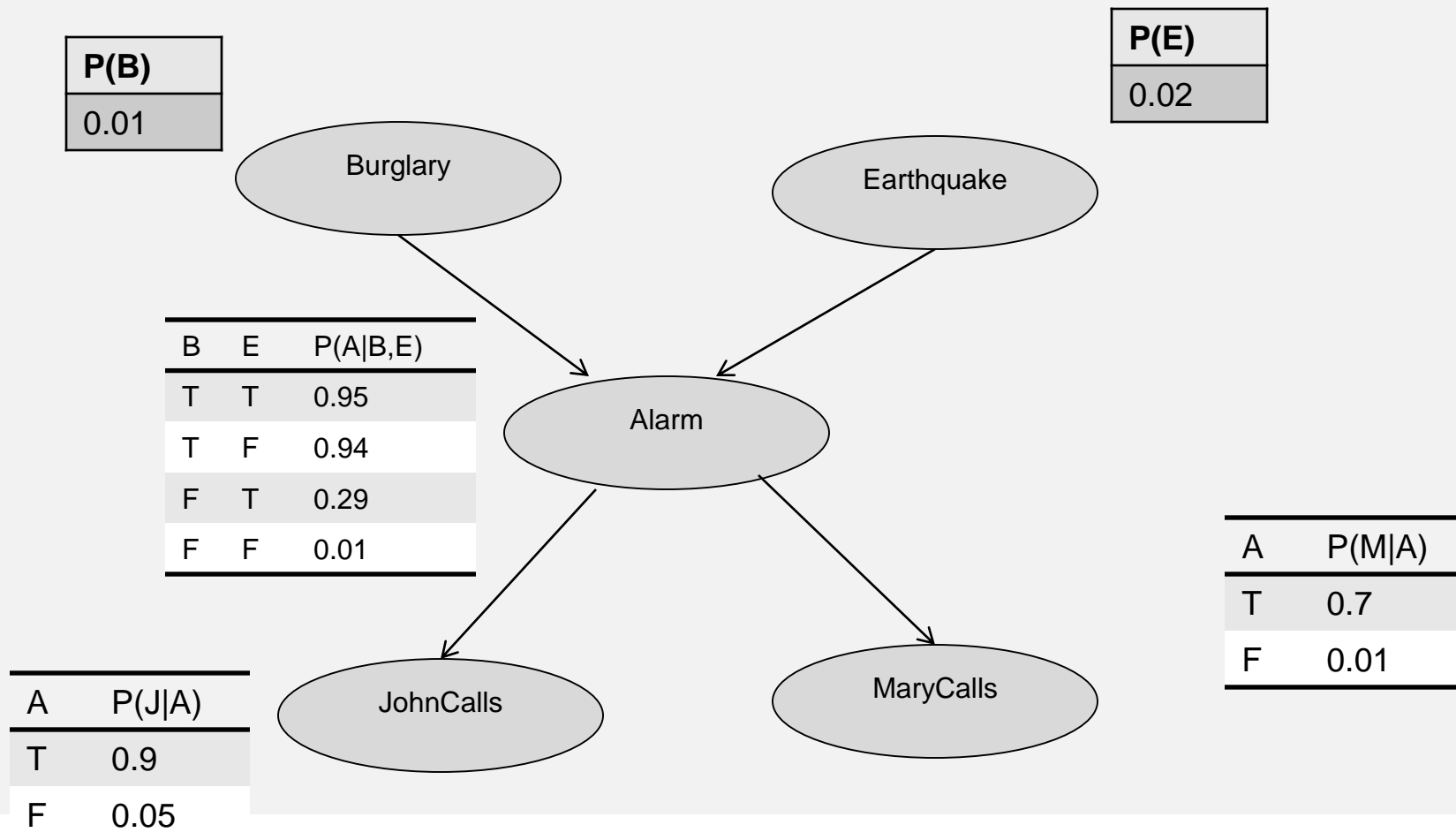
Exemplo



Exemplo



Exemplo



Topologia da Rede

Roubos e terremotos
afectam diretamente a
probabilidade do
alarme tocar

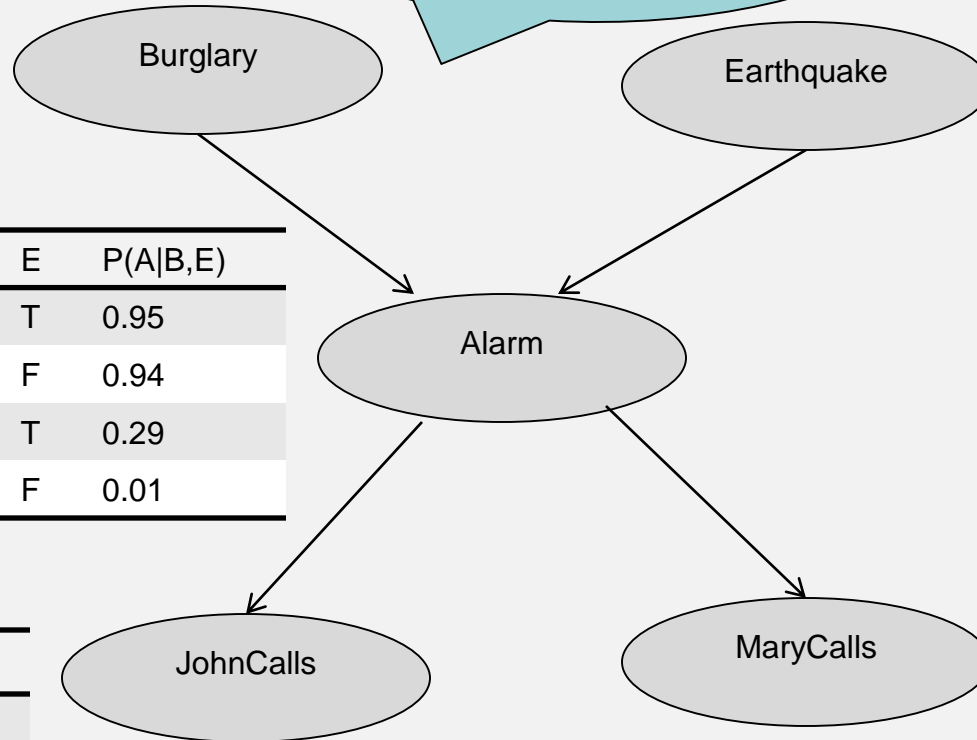
$P(B)$
0.01

$P(E)$
0.02

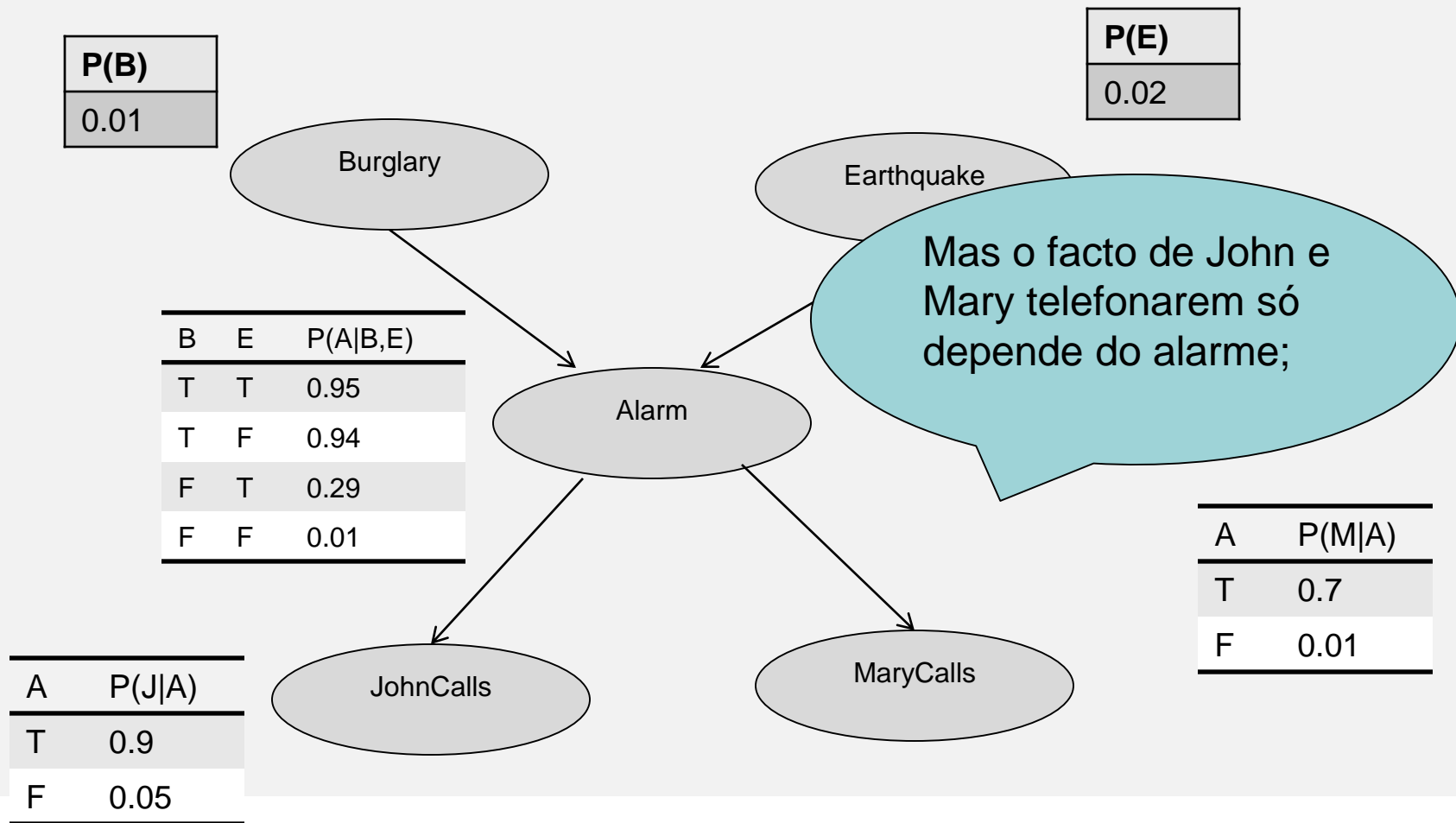
B	E	$P(A B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.01

A	$P(J A)$
T	0.9
F	0.05

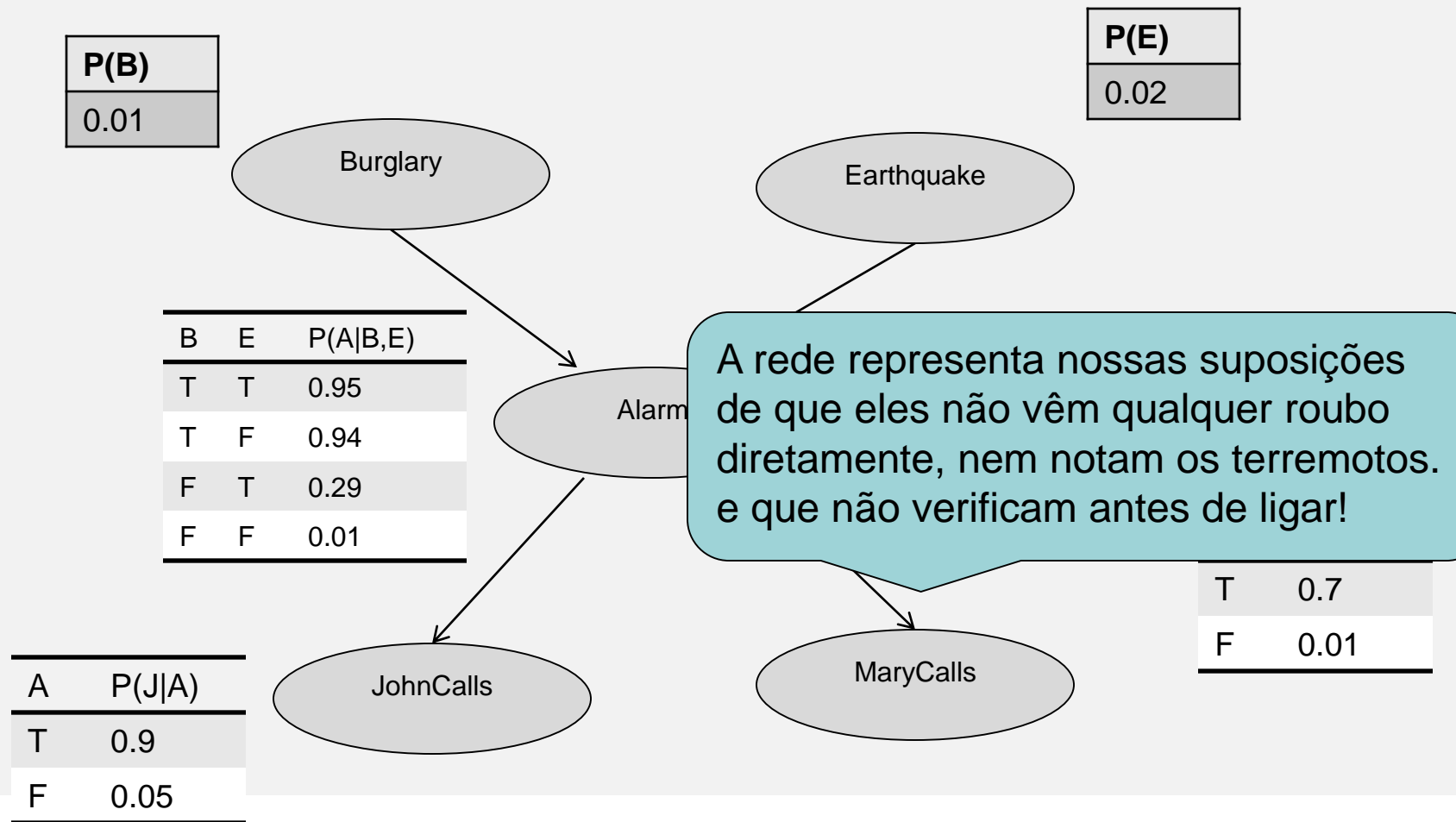
A	$P(M A)$
T	0.7
F	0.01



Topologia da Rede



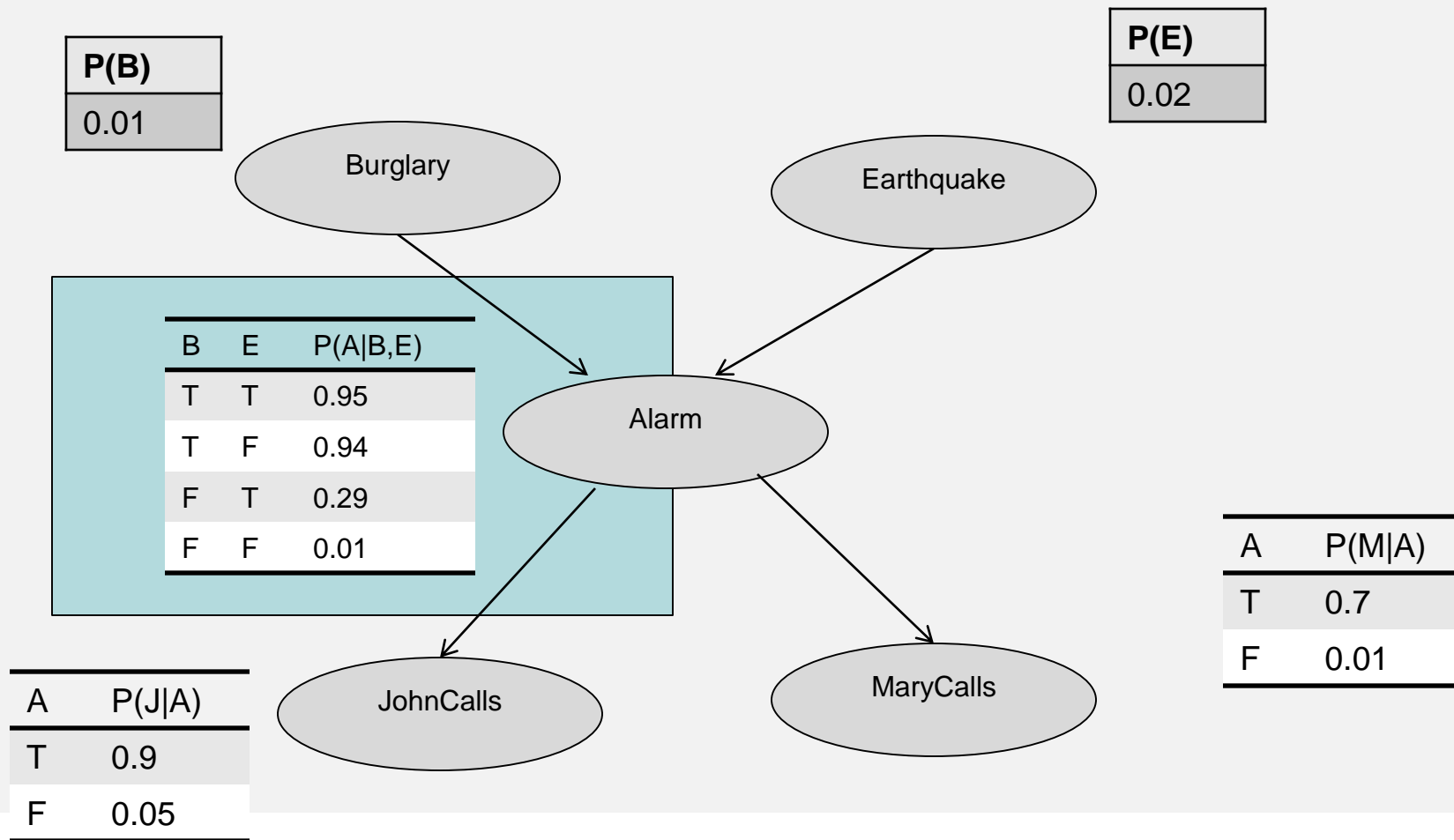
Topologia da Rede



As probabilidades...

- ... resumem um conjunto potencialmente infinito de circunstâncias (Mary ouve música alta, John liga quando toca o telefone; humidade, falta de energia, etc podem interferir no alarme;) mas não temos que as listar...
- **Preguiça e ignorância!**

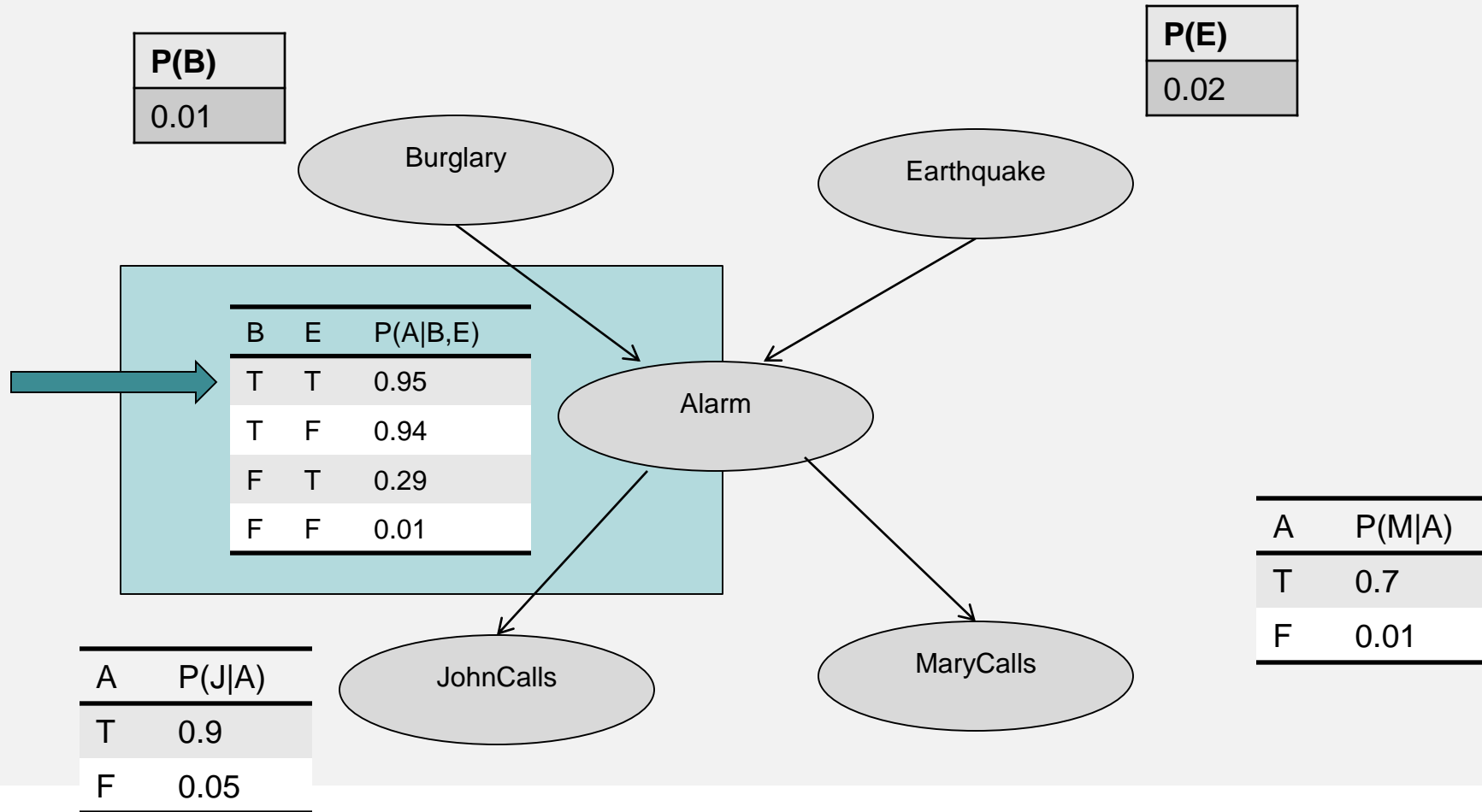
Exemplo



Tabelas de probabilidade condicional (TPC)

- Cada linha em uma TPC contém a probabilidade condicional de cada valor de nó para um **caso de condicionamento**;
 - um caso de condicionamento é apenas uma combinação possível de valores para os nós superiores

Tabelas de probabilidade condicional (TPC)



Tabelas de probabilidade condicional (TPC)

- Cada linha em uma TPC contém a probabilidade condicional de cada valor de nó para um **caso de condicionamento**;
 - um caso de condicionamento é apenas uma combinação possível de valores para os nós superiores
- Para cada nó X_i da rede, cada linha da TPC requer um número p para $X_i = \text{true}$ (prob. para $X_i = \text{false}$ é $1-p$)

Tabelas de probabilidade condicional (TPC)

- Cada linha em uma TPC contém a probabilidade condicional de cada valor de nó para um **caso de condicionamento**;
 - um caso de condicionamento é apenas uma combinação possível de valores para os nós superiores
- Para cada nó X_i da rede, cada linha da TPC requer um número p para $X_i = \text{true}$ (prob. para $X_i = \text{false}$ é $1-p$)
- Um nó sem pais tem apenas uma linha: “probabilidade a priori”
- Em geral, uma tabela para uma variável booleana com k pais booleanos possui 2^k probabilidades
- Se cada variável não possuir mais do que k pais, uma consulta à rede completa será $O(n \cdot 2^k)$, para n = número de nós.
 - I.e., cresce linearmente em n , vs. $O(2^n)$ da distribuição total

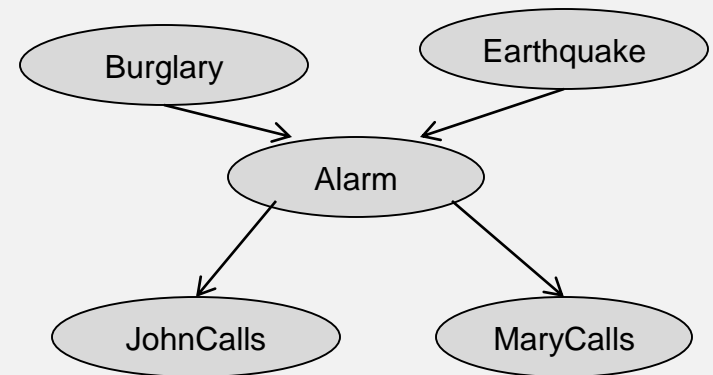
Semântica das Redes Bayesianas

- Duas maneiras equivalentes:
 - **Semântica global** (ou numérica): entender as redes como uma representação da distribuição de probabilidade conjunta;
 - indica como construir uma rede
 - **Semântica local** (ou topológica): visualizá-las como uma codificação de uma colecção de declarações de independência condicional.
 - Indica como fazer inferências com uma rede.

Semântica Global

A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:

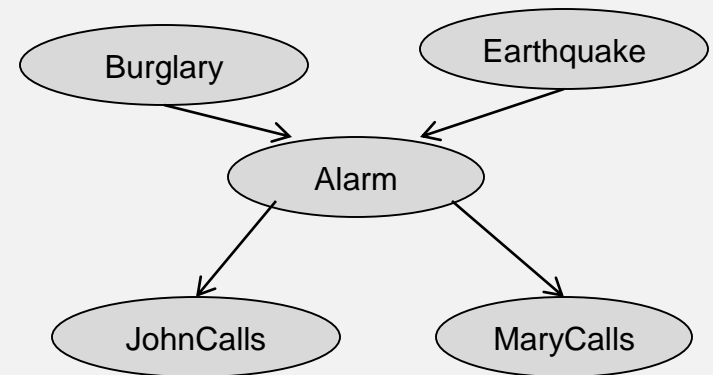
$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1} \mathbf{P}(X_i | \text{Parents}(X_i))$$



Semântica Global

Se uma rede bayesiana for uma representação da distribuição conjunta, então ela poderá ser utilizada para responder a qualquer consulta efectuando-se o somatório de todas as entradas conjuntas relevantes

B	E	$P(A B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.01



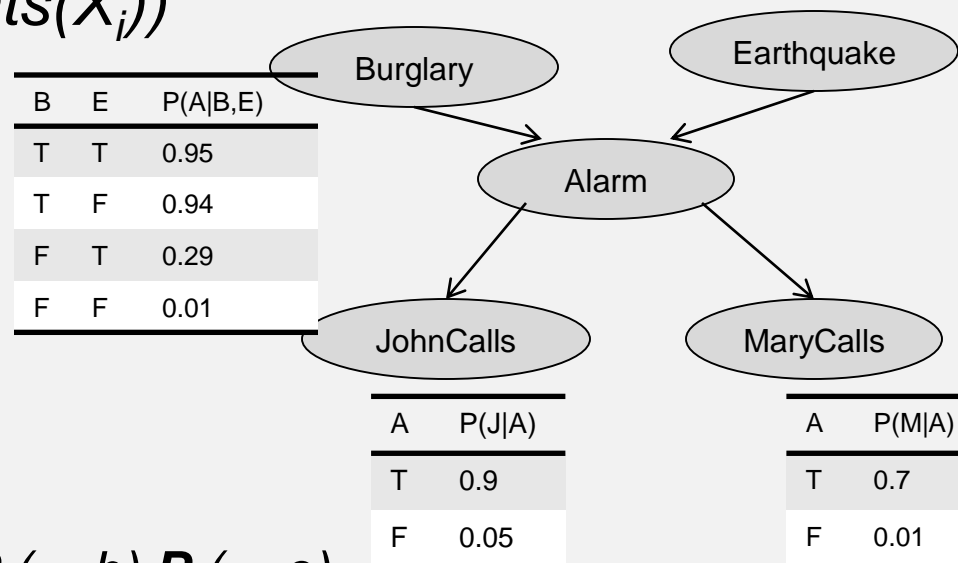
A	$P(J A)$
T	0.9
F	0.05

A	$P(M A)$
T	0.7
F	0.01

Exemplo

A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i))$$



$$P(j, m, a, \neg b, \neg e) = \text{????}$$

$$= P(j | a) P(m | a) P(a | \neg b, \neg e) P(\neg b) P(\neg e)$$

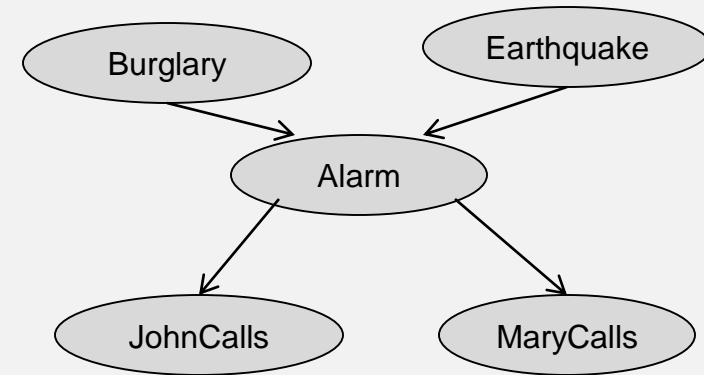
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$

$$= 0.00063$$

Semântica local

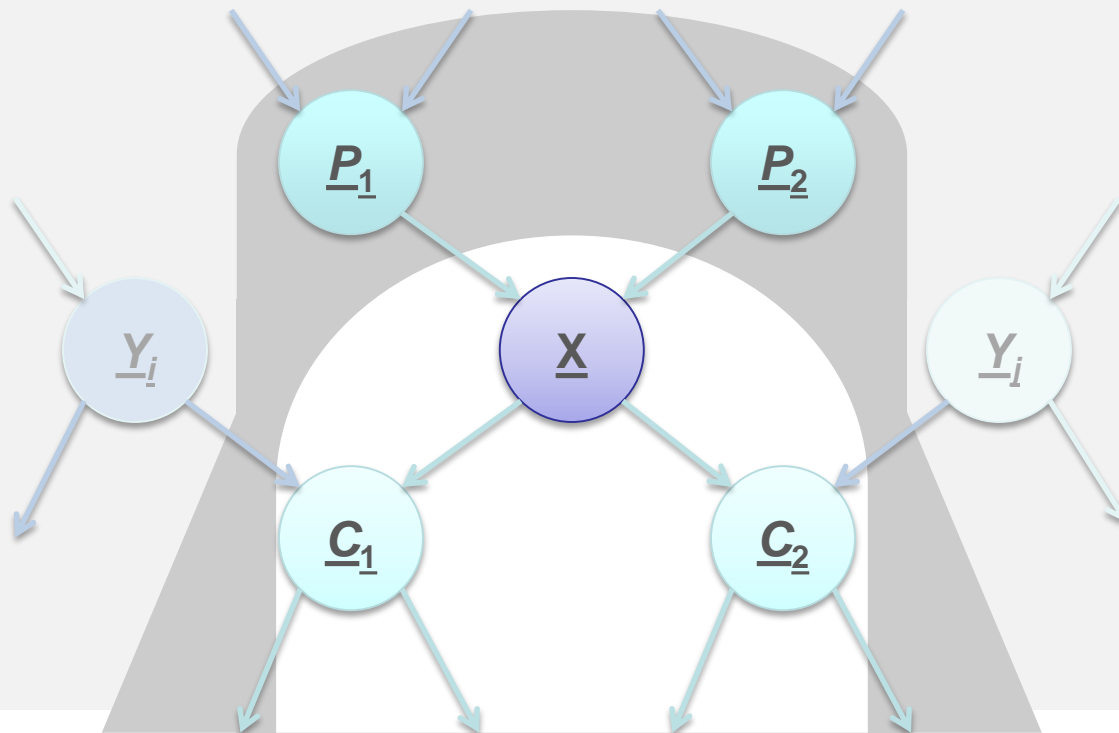
Semântica local (topológica):
cada nó é condicionalmente independente de seus não-descendentes, dados seus pais.

- Ex. **JohnCalls** deve ser independente de **Burglary** e **Earthquake** dado **Alarm**!



Semântica Local

Cada nó é condicionalmente independente dos seus não descendentes dado os seus pais



Método para construção de redes Bayesianas



Semântica Global

A semântica global (ou numérica) define a distribuição de probabilidade total como o produto das distribuições condicionais locais:

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

Ou seja:

$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | \text{Parents}(X_i))$$

Chain Rule

Construção de redes Bayesianas

- A partir das condições de independência dada pela semântica local

- $P(X_1, \dots, X_n) = \prod_{i=1} P(X_i | \text{Parents}(X_i))$

- A distribuição conjunta total pode ser obtida (lembrando da regra do produto):

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

- $P(X_1, \dots, X_n) = \prod_{i=1} P(X_i | X_{i-1}, X_{i-2}, \dots, X_1)$ **Regra da cadeia**

Construção de redes Bayesianas

- De $\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | \text{Parents}(X_i))$ e a regra da cadeia:
 - $\mathbf{P}(X_i | \text{Pais}(X_i)) = \mathbf{P}(X_i | X_1, \dots, X_{i-1})$
- a rede bayesiana é uma representação correta do domínio **somente se cada nó for condicionalmente independente de seus predecessores na ordenação dos nós**, dados seus pais.
- Ou seja:
 - precisamos escolher pais para cada nó de forma que essa propriedade se mantenha
 - I.e. os pais de X_i devem conter todos os nós $X_{i-1}, X_{i-2}, \dots, X_1$ que o influenciam directamente.

Construção de uma rede Bayesiana

- 1. Escolher uma ordem para as variáveis aleatórias X_1, \dots, X_n
- 2. Para $i = 1$ à n
 - adicione X_i à rede
 - selecione pais para X_1, \dots, X_{i-1} tais que
$$\mathbf{P}(X_i \mid \text{Pais}(X_i)) = \mathbf{P}(X_i \mid X_1, \dots, X_{i-1})$$

Esta escolha de pais garante a semântica global:

$$\begin{aligned}\mathbf{P}(X_1, \dots, X_n) &= \prod_{i=1} \mathbf{P}(X_i \mid X_1, \dots, X_{i-1}) \text{ (regra da cadeia)} \\ &= \prod_{i=1} \mathbf{P}(X_i \mid \text{Pais}(X_i)) \text{ (por construção)}\end{aligned}$$

Ordem para as variáveis

- A ordem correta em que os nós devem ser adicionados consiste em adicionar primeiro as “**causas de raiz**”, depois as variáveis que elas influenciam e assim por diante, até chegarmos às folhas, que não têm nenhuma influência causal directa sobre as outras variáveis.
- E se escolhermos a ordem “errada”??

Ordem para as variáveis

- A ordem correta em que os nós devem ser adicionados consiste em adicionar primeiro as “**causas de raiz**” (as variáveis que não dependem de outras), depois as variáveis que elas influenciam e assim por diante, até chegarmos às folhas, que não têm nenhuma influência causal directa sobre as outras variáveis.
- E se escolhermos a ordem “errada”??

Intuitivamente, os pais de um nó X_i deve conter todos os nós X_1, \dots, X_{i-1} que influenciem directamente X_i

Exemplo

- Assumindo a ordem: M, J, A, B, E

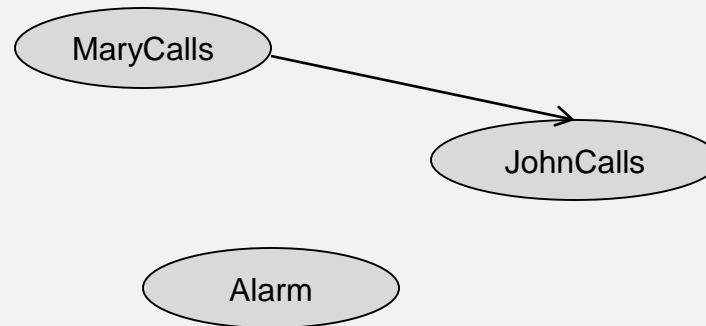


$$P(J \mid M) = P(J)?$$

*Se a Mary ligar, quer dizer que o alarme disparou o que significa que é mais provavel que o John também ligue..
Ou seja JohnCalls tem que ter a MaryCalls como pai*

Exemplo

- Assumindo a ordem: M, J, A, B, E

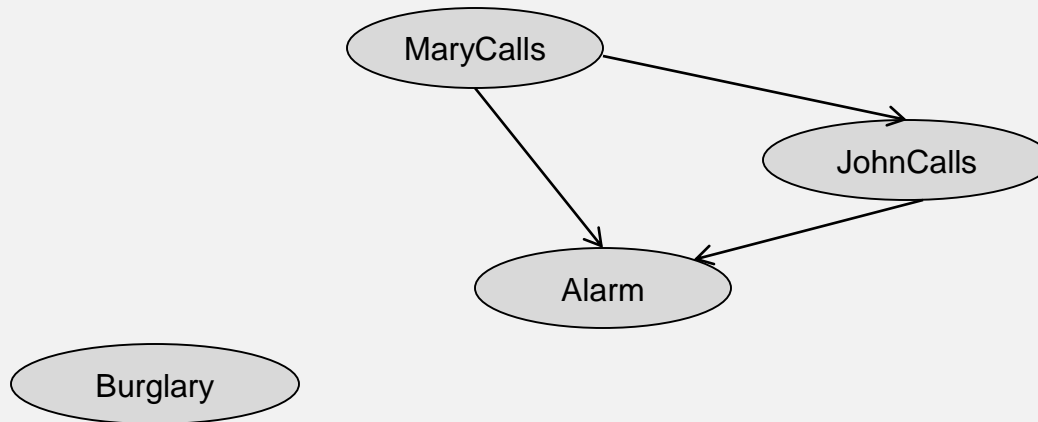


$P(J \mid M) = P(J)$? No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?

Exemplo

- Assumindo a ordem: M, J, A, B, E



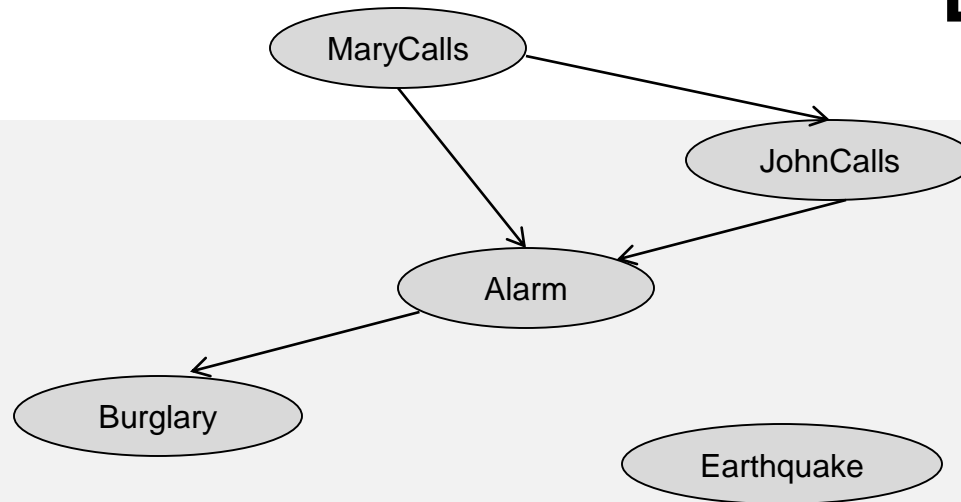
$P(J \mid M) = P(J)$? No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No

$P(B \mid A, J, M) = P(B \mid A)$?

$P(B \mid A, J, M) = P(B)$?

Exemplo



$P(J \mid M) = P(J)$? No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No

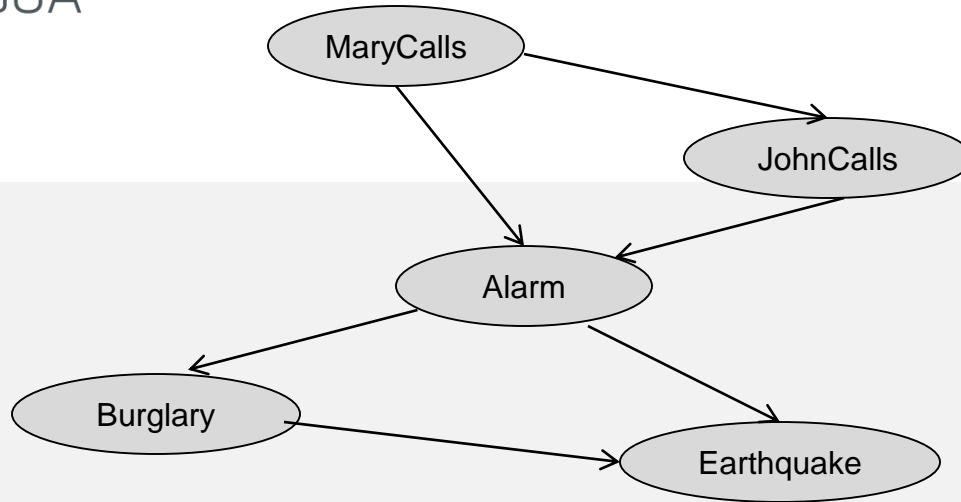
$P(B \mid A, J, M) = P(B \mid A)$? Yes

$P(B \mid A, J, M) = P(B)$? No

$P(E \mid B, A, J, M) = P(E \mid A)$?

$P(E \mid B, A, J, M) = P(E \mid A, B)$?

Exemplo



$P(J \mid M) = P(J)$? No

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No

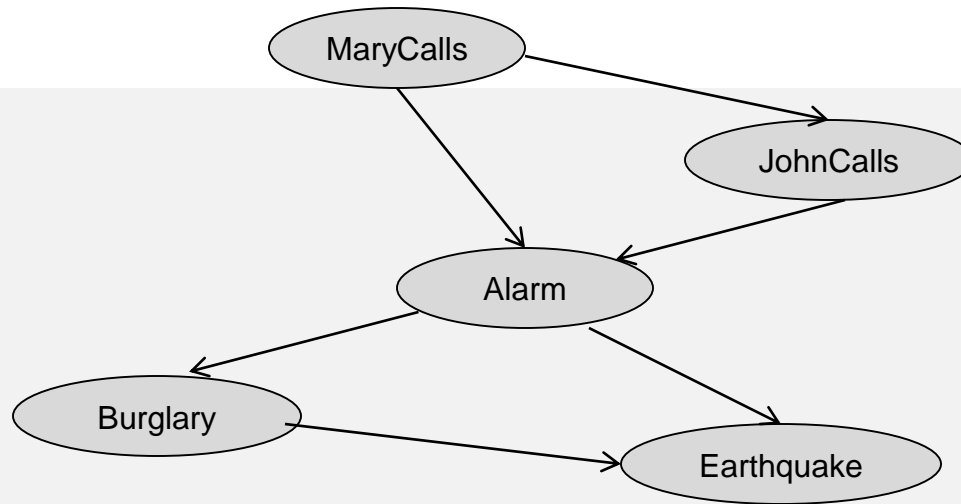
$P(B \mid A, J, M) = P(B \mid A)$? Yes

$P(B \mid A, J, M) = P(B)$? No

$P(E \mid B, A, J, M) = P(E \mid A)$? No

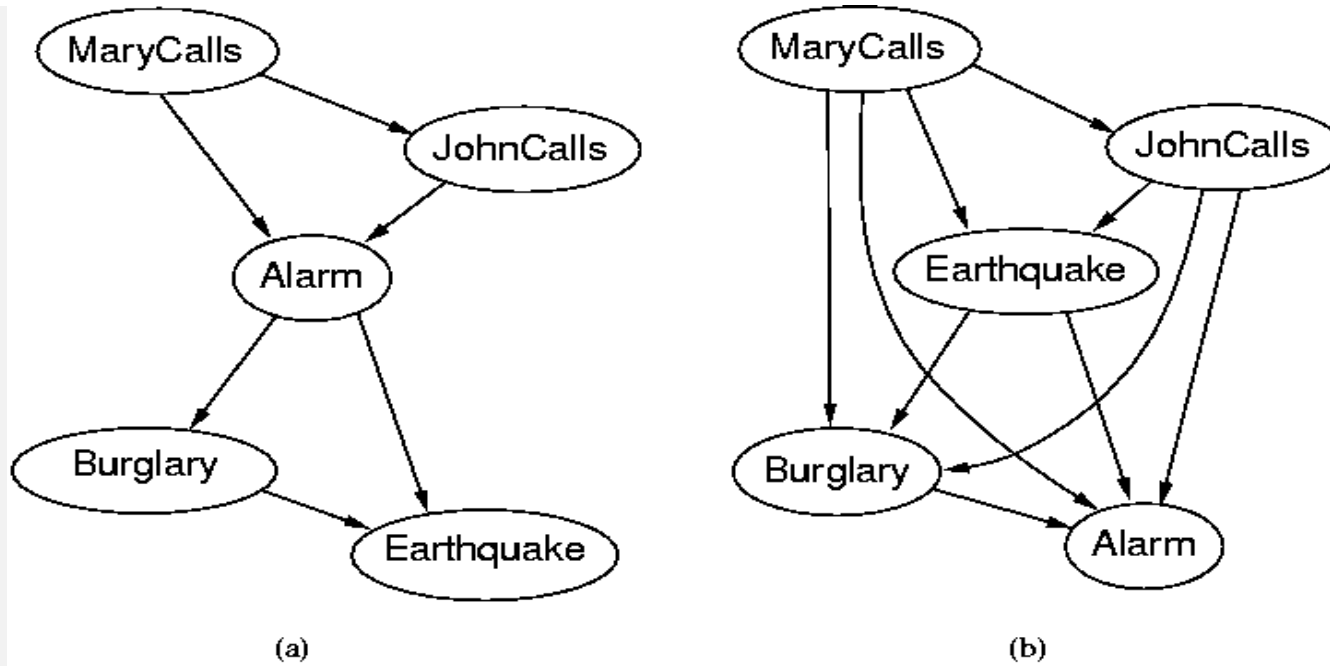
$P(E \mid B, A, J, M) = P(E \mid A, B)$? Yes

Exemplo



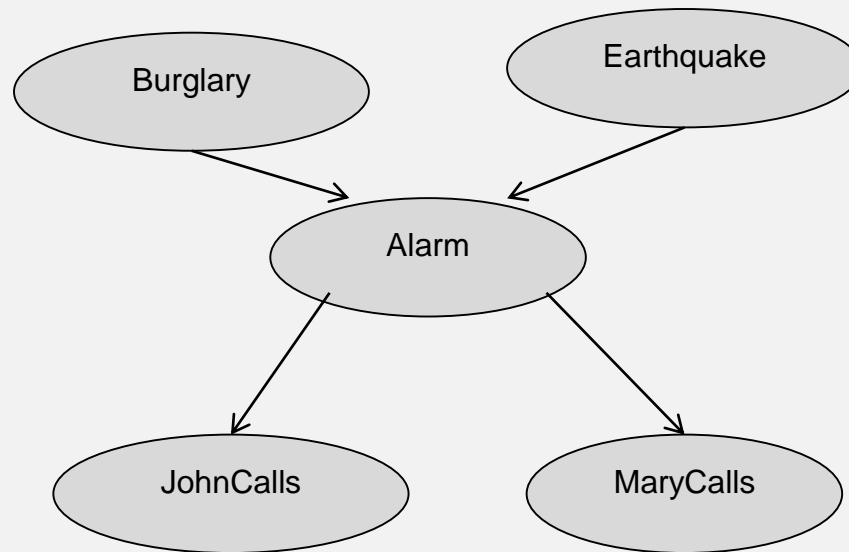
- A rede resultante terá dois vínculos a mais que a rede original e **exigirá outras probabilidades para serem especificadas**
- Alguns dos vínculos apresentam relacionamentos tênues que exigem julgamentos de probabilidade difíceis e antinaturais (**prob de *Terremoto*, dados *Roubo* e *Alarme***)
- (Em geral) é melhor pensar **de causas para efeitos** (modelo causal) e não do contrário (modelo de diagnóstico)

Exemplo



- Uma ordenação de nós má: *MarryCalls*, *JohnCalls*, *Earthquake*, *Burglary* e *Alarm*
- Entretanto, todas as três redes devem representar a **mesma distribuição conjunta**. As duas últimas só não expressam todas as independências condicionais

Rede final: boa ordenação



Representação eficiente de distribuições condicionais

- Ainda que o número de pais k seja reduzido, o preenchimento da TPC para um nó exige até $O(2^k)$ e muita experiência para decidir os casos condicionais.
 - Esse é o pior caso, em que os relacionamentos de pais e filhos é arbitrário
- Em muitos casos podemos utilizar um padrão para obter a tabela (distribuição canónica).

Representação eficiente de distribuições condicionais

Em muitos casos podemos utilizar um “padrão” para obter a tabela (distribuição canónica).

Representação eficiente de distribuições condicionais

Distribuição canónica:

- ajustar a distribuição de probabilidades em cada nó a alguma forma padrão;
- nestes casos a tabela completa pode ser especificada nomeando-se o padrão e fornecendo-se alguns parâmetros.
- Exemplos:
 - nós determinísticos
 - relacionamentos lógicos ruidosos: *ou-ruidoso*

Representação eficiente: Distribuição canónica

- **Nós determinísticos:** tem seus valores especificados pelos valores de seus pais, sem qualquer incerteza:
 - $X = f(\text{Pais}(X))$ para alguma função f ;
 - funções booleanas:
 - Norte Americano = Canadense \vee US \vee Mexicano
 - relação numérica entre funções contínuas:
 - *Valor mínimo de alguma função*

Relações Incertas: Noisy-OR



Caso: fever *iff* (cold or flu or malaria)

Relações Incertas: Noisy-OR

- Noisy-OR (ou “Ou-ruidoso” em contraste com o “Ou” proposicional)
 - Permite a incerteza sobre a capacidade (força) de cada pai para fazer o filho ser verdadeiro
 - Relacionamento entre pai e filho pode ser inibido.
 - Inibições independentes
 - Assim “febre” é falsa sse todos os seus “pais verdadeiros são inibidos”, e a probabilidade de isso ocorrer é o produto das probabilidades de inibição de cada pai
 - $P(X_i | Pais(X_i)) = 1 - \prod q_j$

Relações Incertas: Noisy-OR

Probabilidades de inibição:

$$Q_{\text{cold}} = P(\neg \text{fever} \mid \text{cold}, \neg \text{flu}, \neg \text{malaria}) = 0.6$$

$$Q_{\text{flu}} = P(\neg \text{fever} \mid \neg \text{cold}, \text{flu}, \neg \text{malaria}) = 0.2$$

$$Q_{\text{malaria}} = P(\neg \text{fever} \mid \neg \text{cold}, \neg \text{flu}, \text{malaria}) = 0.1$$

<i>Cold</i>	<i>Flu</i>	<i>Malaria</i>	<i>P(Fever)</i>	<i>P(¬Fever)</i>
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Relações Incertas: Noisy-OR

Em geral, as relações incertas em que a variável depende de k pais pode ser descrita usando $O(k)$ parâmetros face aos $O(2^k)$ necessários para a tabela de probabilidade condicional total.

Outros tipos de Redes

- Muitos problemas do mundo real envolvem variáveis com valores contínuos (eg. altura, volume, temperatura, dinheiro...).
- Uma forma de lidar com variáveis contínuas é através de discretização, isto é dividir os valores em intervalos fixos. Eg. Temperatura: dividida em ($<0^{\circ}\text{C}$), ($0^{\circ}\text{C}-100^{\circ}\text{C}$) e ($>100^{\circ}\text{C}$)
- Embora a discretização seja uma possível solução, tem problemas de perda de precisão.
- A forma de resolver é definir famílias standard de funções de distribuição, por exemplo uma distribuição Gaussiana, $N(\mu, \sigma^2)$

Redes de Bayes Híbridas

Uma Rede Bayesiana Híbrida é uma rede que contém simultaneamente variáveis discretas e contínuas.

- Para especificar uma rede Bayesiana híbrida temos que especificar os dois tipos de distribuição:
 - Distribuição condicional para uma variável contínua dados pais discretos
 - Distribuição condicional para uma variável discreta dados pais contínuos

Mas para que servem as
redes Bayesianas?

Para descobrir relações é necessário fazer
“Inferência”...

Inferência exacta em Redes Bayesianas

A tarefa essencial para um sistema de inferência com redes Bayesianas é a determinação da distribuição de probabilidades posteriores para um conjunto de variáveis dado um evento observado.

Tipos de Inferência

Inferência de Diagnóstico

Dos efeitos para as causas

- E.g. Dado que o John ligou, inferir a probabilidade de ser um roubo
 $P(\text{Burglary}|\text{JohnCalls})$

Tipos de Inferência

Inferência Causal

Das causas para os efeitos:

Ex: dado que houve um roubo, qual a probabilidade de John e a Mary ligarem
 $P(\text{JohnCalls}|\text{Burglary})$ and
 $P(\text{MaryCalls}|\text{Burglary})$

Tipos de Inferência

Inferências “inter-causais”

- “dado que o alarme disparou...”, nós temos que $P(\text{Burglary}|\text{Alarm}) = 0.376$.
... mas.... “se adicionarmos a evidência que houve um tremor de terra”, então $P(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake})$ desce para 0.003....

Inferência Exacta

- $P(X|e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$
- X - variável que estamos a investigar
- e - conjunto de variáveis que constitui a “evidencia”, i.e. e_1, e_2, \dots, e_m
- y – variáveis que não são evidência, mas estão escondidas (“Hidden variables”)
- $\alpha = \frac{1}{P(e)}$

Tipos de Inferência: combinação

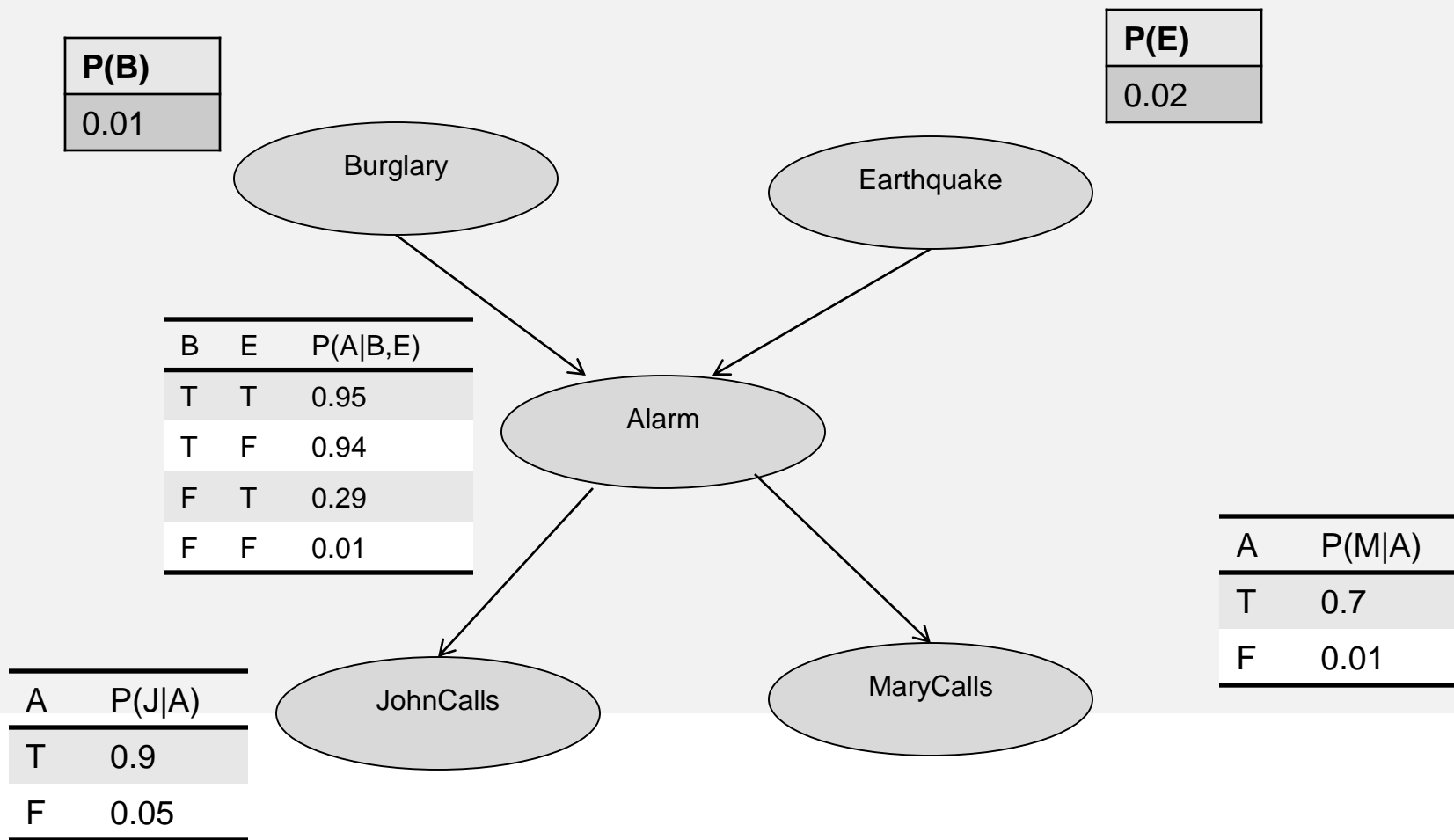
Por exemplo: Qual a probabilidade de ter um ladrão se a Mary e o John ambos me ligarem?

Query: é obtida pelo cálculo da distribuição de probabilidades posterior $P(X|e)$
 $P(\text{Burglar}|\text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true})$ ou, $P(B|j, m)$

Inferência por enumeração

Vamos ao exemplo:

$$P(b|j,m) = \text{?????}$$



Inferência por enumeração

Uma forma inteligente de somar as variáveis da distribuição sem construir a sua representação explícita

$P(\text{Burglary} \mid \text{JohnCalls}=\text{true}, \text{MaryCalls}=\text{true})$

$$\begin{aligned} P(b \mid j, m) \\ &= \alpha P(b, j, m) \\ &= \alpha \sum_e \sum_a P(b, e, a, j, m) \end{aligned}$$

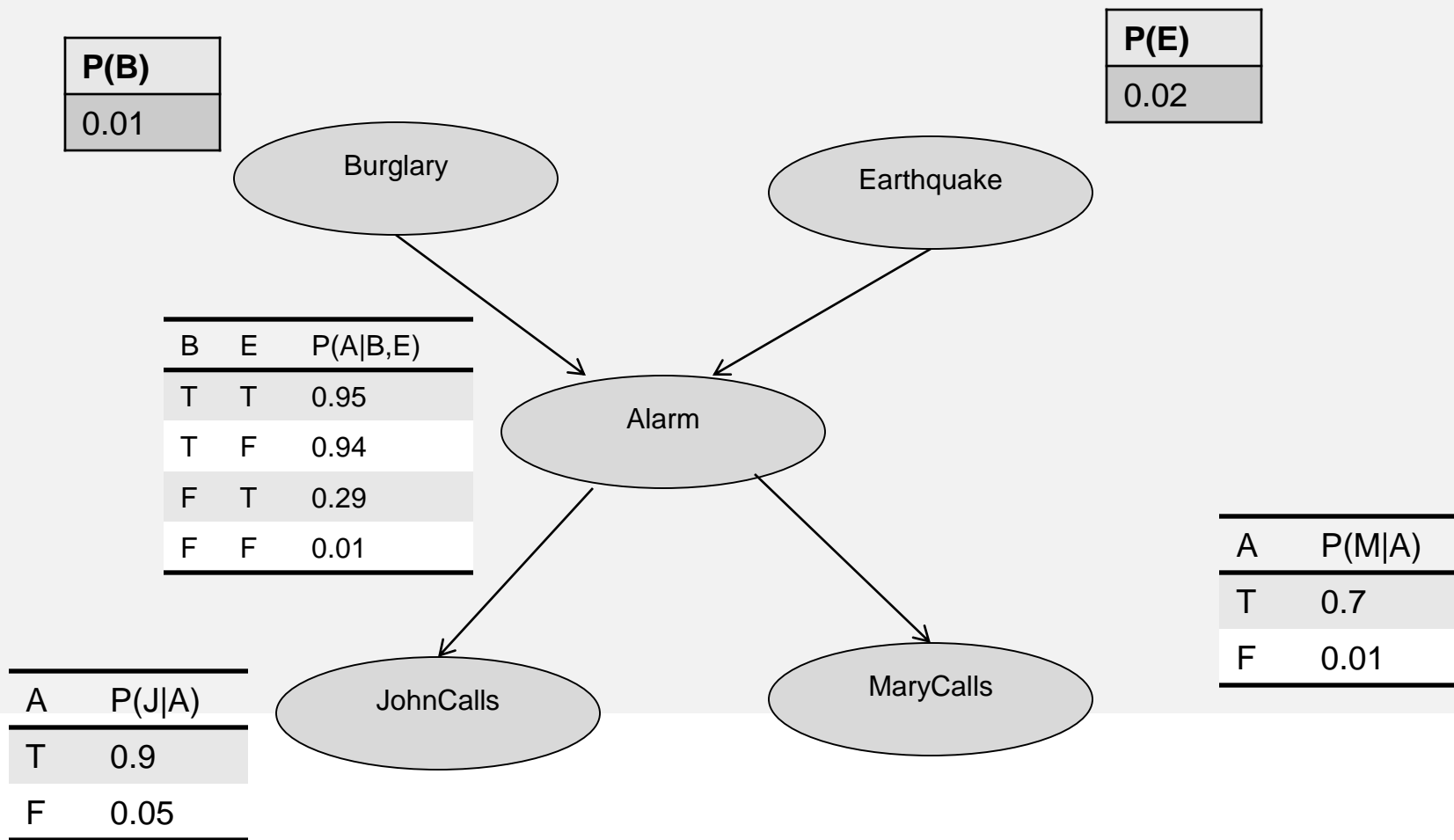
Em que “**e**” e “**a**” são as variáveis escondidas

$$\begin{aligned} P(b \mid j, m) \\ &= \alpha \sum_e \sum_a P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a) \\ &= \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \end{aligned}$$

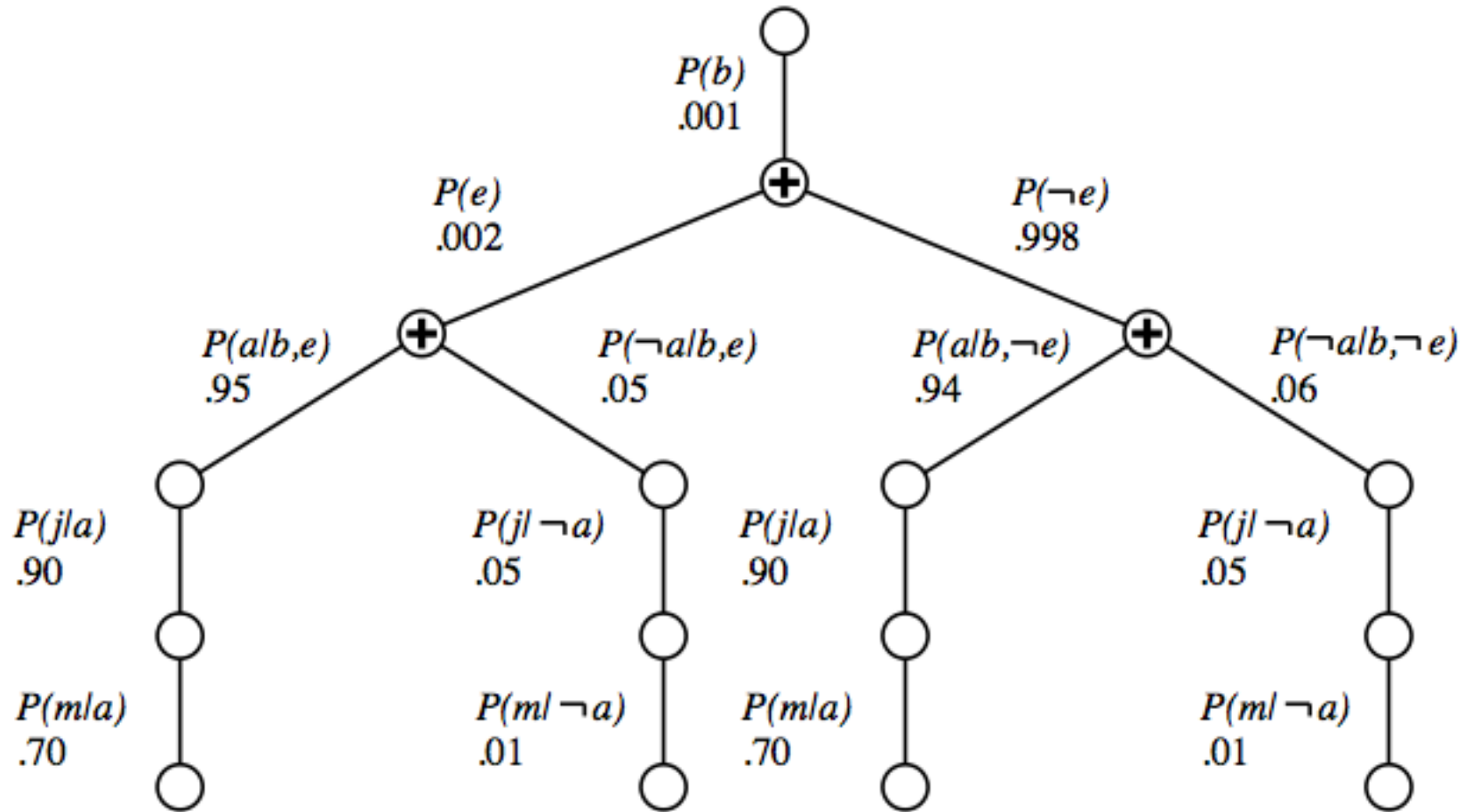
Inferência por enumeração

Vamos ao exemplo:

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



Árvore de Avaliação



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

function ENUMERATION-ASK(X, e, bn) **returns** a distribution over X

inputs: X , the query variable

e , observed values for variables E

bn , a Bayes net with variables $\{X\} \cup E \cup Y$ /* $Y = \text{hidden variables}$ */

$Q(X) \leftarrow$ a distribution over X , initially empty

for each value x_i of X **do**

 extend e with value x_i for X

$Q(x_i) \leftarrow$ ENUMERATE-ALL(VARS[bn], e)

return NORMALIZE($Q(X)$)

function ENUMERATE-ALL($vars, e$) **returns** a real number

if EMPTY?($vars$) **then return** 1.0

$Y \leftarrow$ FIRST($vars$)

if Y has value y in e

then return $P(y \mid \text{parents}(Y)) \times \text{ENUMERATE-ALL}(\text{REST}(\text{vars}), e)$

else return $\sum_y P(y \mid \text{parents}(Y)) \times \text{ENUMERATE-ALL}(\text{REST}(\text{vars}), e_y)$
 where e_y is e extended with $Y = y$

Figure 14.9 The enumeration algorithm for answering queries on Bayesian networks.

Inferência por enumeração

- $P(b|j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$
- $P(b|j, m) = \alpha 0.00059224 \approx 0.284$

... a chance de ser um roubo, dado que houve uma chamada dos dois vizinhos é de 28%

Inferência por enumeração

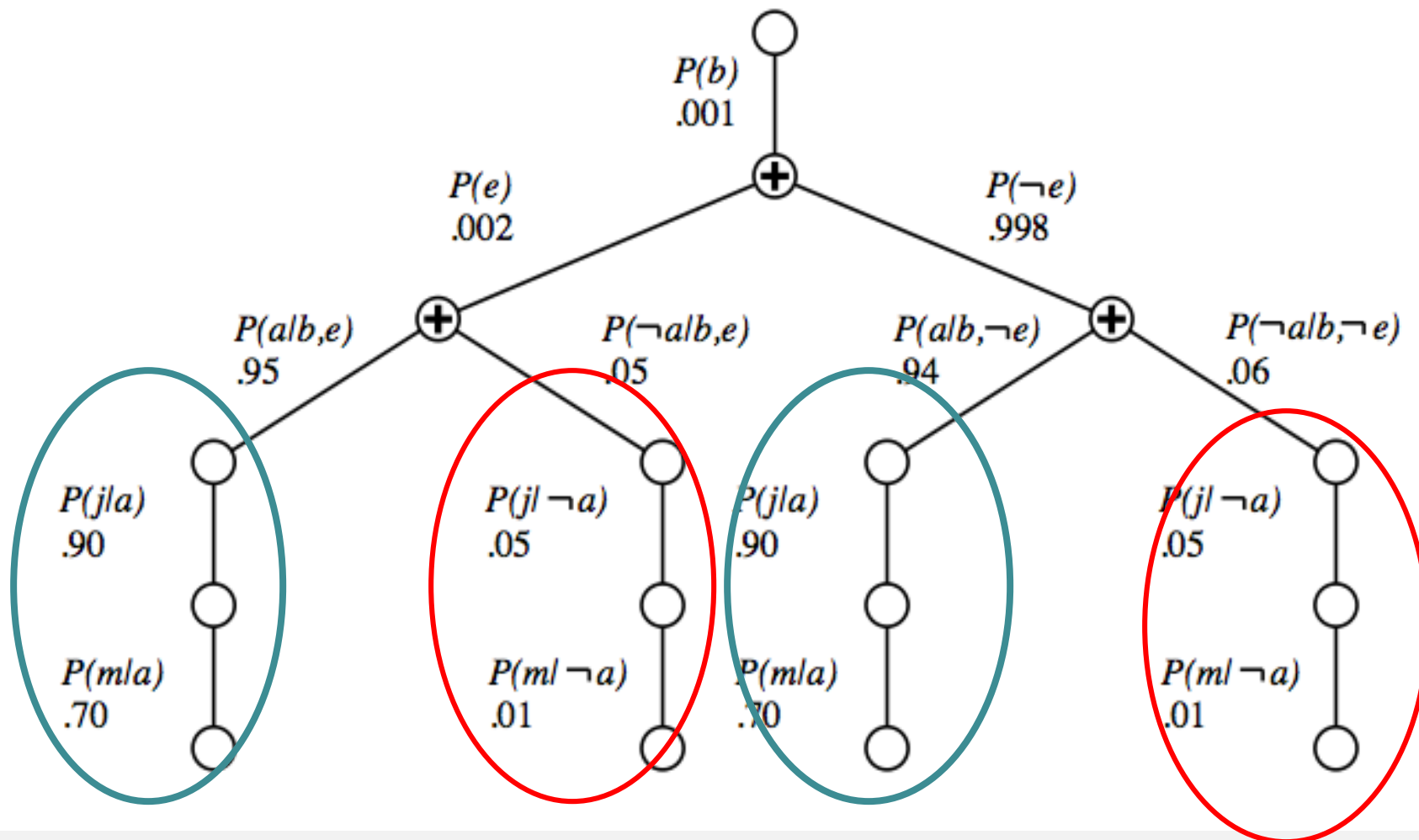
- $P(b|j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$
- $P(b|j, m) = \alpha 0.00059224 \approx \langle 0.284, 0716 \rangle$

A complexidade espacial do algoritmo

ENUMERATION-ASK é linear no número de variáveis.

Infelizmente a complexidade temporal para uma rede com n variáveis é sempre $O(2^n)$

Árvore de Avaliação



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

Eliminação de Variáveis

- A eliminação de variáveis:
 - fazer as somas da direita para a esquerda
 - Ou de baixo para cima na árvore
 - guardando os resultados intermédios (factores) para evitar ter que os voltar a calcular...

$$\begin{aligned} \mathbf{P}(B|j, m) &= \alpha \mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B, e) P(j|a) P(m|a) \\ &= \alpha \mathbf{f}_1(B) \sum_e \mathbf{f}_2(E) \sum_a \mathbf{f}_3(A, B, E) \mathbf{f}_4(A) \mathbf{f}_5(A) \end{aligned}$$

Em que $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ e \mathbf{f}_5 são vectores de dois elementos, e \mathbf{f}_3 é uma matriz $2 \times 2 \times 2$

Eliminação de Variáveis

$$P(b|j,m) \propto \underbrace{P(b)}_{f_1(B)} \underbrace{\sum_e P(e)}_{f_2(E)} \underbrace{\sum_a P(a|b, e)}_{f_3(A,B,E)} \underbrace{P(j|a)}_{f_5(A)} \underbrace{P(m|a)}_{f_4(A)}$$

$$f_4(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix} \quad f_5(A) = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}$$

Eliminação de Variáveis

$$P(b|j,m) \propto \underbrace{P(b)}_{f_1(B)} \underbrace{\sum_e P(e)}_{f_2(E)} \underbrace{\sum_a P(a|b, e)}_{f_3(A,B,E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)}$$

$$f_5(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix} \quad f_4(A) = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}$$

Calculo dos produtos

$$\begin{aligned} f_6(B, E) &= \sum_a f_3(A, B, E) \times f_4(A) \times f_5(A) \\ &= f_3(a, B, E) \times f_4(a) \times f_5(a) + f_3(\neg a, B, E) \times f_4(\neg a) \times f_5(\neg a) \\ f_7(B) &= \sum_e f_2(E) \times f_6(B, E) = f_2(e) \times f_6(B, e) + f_2(\neg e) \times f_6(B, \neg e) \end{aligned}$$

Finally, we get this:

$$P(B|j, m) = \alpha f_1(B) \times f_7(B)$$

Variáveis irrelevantes

Por exemplo na query

$P(\text{JohnCalls} \mid \text{Burglary}=\text{true})$

$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a|b, e) P(J|a) \sum_m P(m|a)$$

Mas a soma em m é igual a 1; ou seja, Mary não é relevante para esta query.

Todas as variáveis que não são antepassadas de uma query, ou que sejam evidência, são irrelevantes para a query

Inferência exata

- A complexidade da Inferência exata depende muito da estrutura da rede.
- A rede apresentada pertence a uma família de redes onde há no máximo um caminho (não direcionado) entre dois nós da rede.
- Estas redes são chamadas “polytrees” (singly connected) e têm propriedades interessantes:
 - A complexidade espacial e temporal da inferência exata é linear no tamanho da rede
- Em redes “multiply connected” a eliminação de variáveis pode ter complexidade exponencial (em tempo e espaço).

Resumo

- Redes Bayesianas são representações explícitas de independência condicional
- Topologia + TPCs = representações compactas de distribuições conjuntas totais
- Ferramentas poderosas para construir uma representação de um domínio que envolva incerteza.