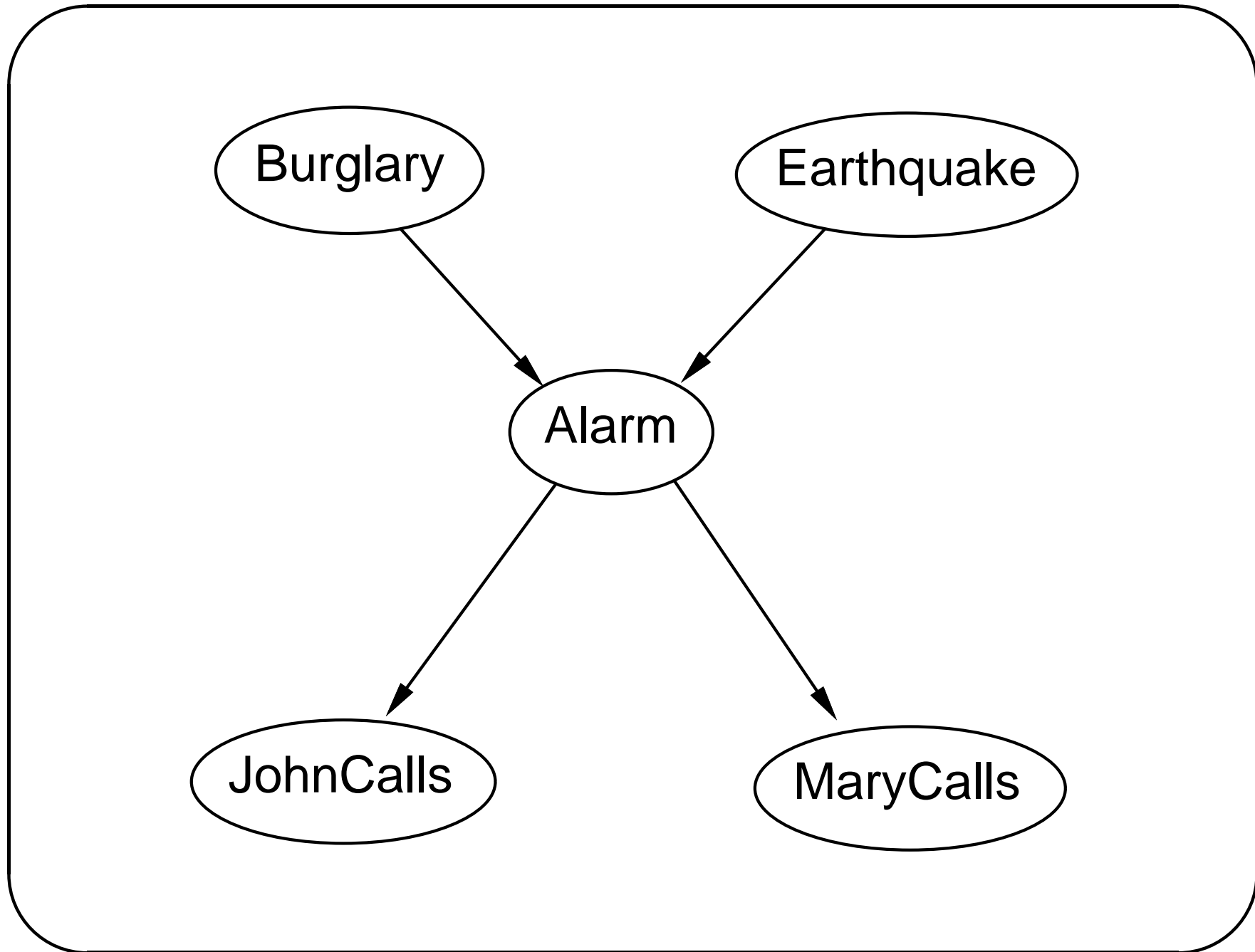


## Cap. 15 (1ed) e 14 (3ed) Sistemas de Raciocínio Probabilístico

- Como construir sistemas de raciocínio utilizando modelos de redes e que usem incerteza de acordo com a teoria das probabilidades?
- **Rede de Crença** ou rede bayesiana: grafo com as seguintes características:
  - Nós representam variáveis aleatórias.
  - Arcos direcionados representam ligações diretas entre variáveis aleatórias.
  - Cada nó tem uma tabela de prob. cond. que quantifica o efeito dos pais deste nó.
  - O grafo não possui ciclos (DAG).
- relativamente fácil para o expert definir as relações do que definir as probs.

## Sistemas de Raciocínio Probabilístico

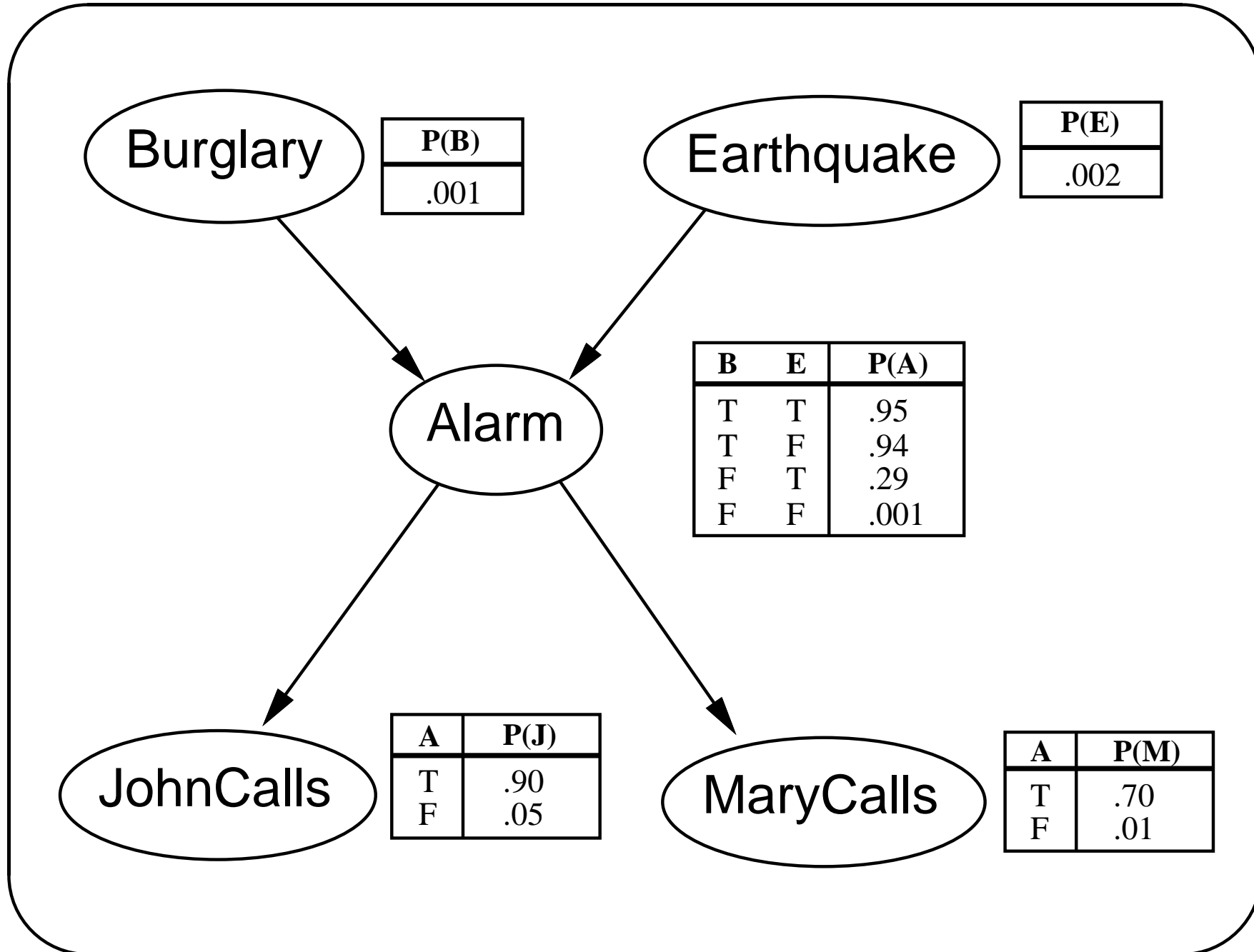
- Exemplo: alarme contra roubo.
- Alarme toca em duas situações: tentativa de roubo e terremoto.
- John e Mary são os vizinhos que avisam ao dono da casa se o alarme estiver tocando.
- John liga toda vez q o alarme toca e tb qdo o tel toca.
- Mary somente liga qdo o alarme toca, mas não ouve algumas vezes.
- Dada a evidência de quem ligou para o dono da casa, queremos descobrir a probabilidade de ter havido roubo.



## Sistemas de Raciocínio Probabilístico

- Rede somente representa ligações diretas, causais.
- Nada é informado sobre Mary ouvir música alta ou de John confundir o tel com o alarme.
- Tabela de probabilidades condicionais:

Roubo	Ter.	$\mathbf{P}(\textit{Alarme} \mid \textit{Roubo}, \textit{Terr})$
T	T	0.950 0.050
T	F	0.950 0.050
F	T	0.290 0.710
F	F	0.001 0.999



## Semântica das Redes de Crenças

- Duas formas de entender:
  - representação da distribuição de probabilidade conjunta. Útil para **construir** redes.
  - conjunto de sentenças condicionalmente independentes. Útil para projetar procedimentos de inferência.
- Representando distribuição de probabilidade conjunta:
  - cada entrada na tabela pode ser calculada através da info na rede.
  - uma entrada genérica representa a probabilidade de uma conjunção de valores atribuídos a cada variável:

$$P(X_1 = x_1 \wedge \dots \wedge X_n = x_n).$$

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid \text{Pais}(X_i))$$

## Construção de Redes de Crenças

- cada entrada na tabela é representada pelo produto dos elementos apropriados das tabelas de probabilidades condicionais (TPCs).
- TPCs fornecem uma representação decomposta da distribuição conjunta.

- Exemplo:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J | A)P(M | A)P(A | \neg B \wedge \neg E)P(\neg B)P(\neg E) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$

- Métodos melhores do que obter a tabela inteira de distribuição de probabilidade conjunta.

## Construção de Redes de Crenças

- Método para construir redes de crenças:
  - Rede é construída de forma que cada nó é condicionalmente independente dos seus predecessores, dada a probabilidade dos seus pais.
  - equação  $P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | Pais(x_i))$  usada para guiar o engenheiro de conhecimento a construir a topologia da rede.
  - Para construir uma rede de forma que esta tenha a estrutura correta para o domínio, escolhe-se nós pais adequados para garantir que cada nó é condicionalmente independente de seus antecessores.



## Construção de Redes de Crenças

- Em geral:

$\mathbf{P}(X_i | X_{i-1}, \dots, X_1) = P(X_i | Pais(X_i))$ , desde que  $Pais(X_i)$  seja subconjunto ou igual a  $\{x_{i-1}, \dots, x_1\}$

- Esta condição é alcançada desde que se rotule os nós da rede em qq ordem q seja consistente com a ordem parcial implícita na estrutura do grafo.
- Ex: Mary telefona: não é **diretamente** influenciado por tentativa de assalto ou terremoto. É influenciado pelo efeito de tentativa de assalto ou terremoto, ou seja, soar o alarme.
- o Fato de John telefonar tb não tem influência direta sobre o fato de Mary telefonar. Neste caso temos independência condicional:

$\mathbf{P}(MaryTel | JohnTel, Alarme, Terr, Assalto) = \mathbf{P}(MaryTel | Alarme)$

## Construção de Redes de Crenças

- Procedimento geral para construção de redes:
  1. Escolha o conj de variáveis  $X_i$  relevantes que descrevam o domínio.
  2. Escolha ordem para as variáveis.
  3. Eqto há vars:
    - a) Pegue uma var  $X_i$  e adicione um nó na rede para  $X_i$ .
    - b) Construir  $Pais(X_i)$  com um conj mínimo de nós que já estejam na rede, tal que a prop de indep cond seja satisfeita.
    - c) Defina a tabela de prob cond  $p/ X_i$ .

## Construção de Redes de Crenças

- Procedimento garante que a rede é acíclica.
- Rede não contém valores de probabilidades redundantes.
- Garante que axiomas da prob não são violados.

## Exemplo

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

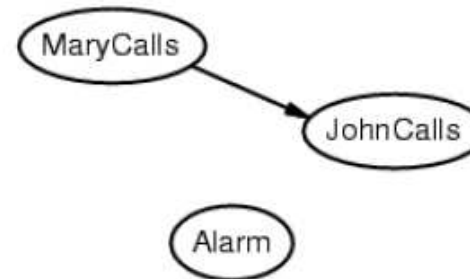
MaryCalls

JohnCalls

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

## Exemplo

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

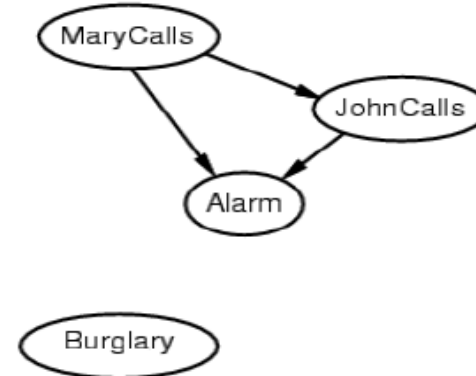


$P(J | M) = P(J)$ ? **No**

$P(A | J, M) = P(A | J)$ ?  $P(A | J, M) = P(A)$ ?

## Exemplo

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



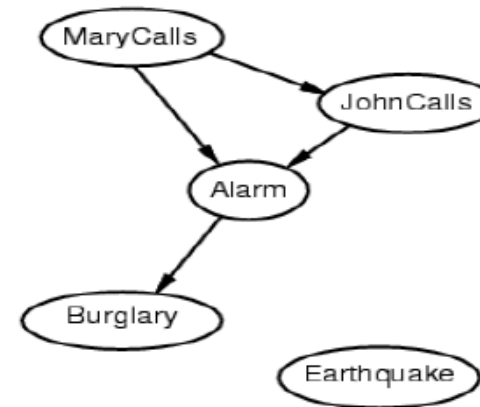
$P(J | M) = P(J)$ ? **No**

$P(A | J, M) = P(A | J)$ ?  $P(A | J, M) = P(A)$ ? **No**

$P(B | A, J, M) = P(B | A)$ ?

$P(B | A, J, M) = P(B)$ ?

## Exemplo



$P(J | M) = P(J)$ ? **No**

$P(A | J, M) = P(A | J)$ ?  $P(A | J, M) = P(A)$ ? **No**

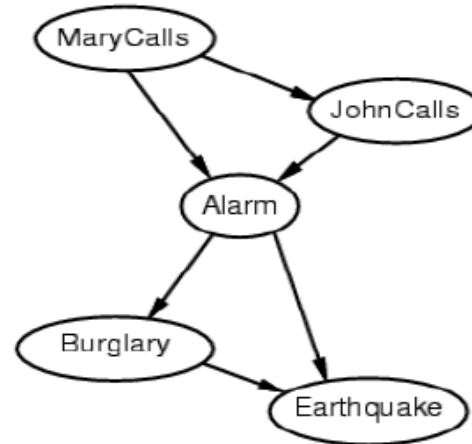
$P(B | A, J, M) = P(B | A)$ ? **Yes**

$P(B | A, J, M) = P(B)$ ? **No**

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

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

## 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**



## Compactação e Ordenação de Nós

- Redes de crenças + **compactas** do que distribuição de prob conjunta.
- Sistemas **localmente estruturados** ou esparsos com info distribuída pelos nós.
- Crescimento polinomial.
- Em redes de crenças, podemos assumir que para a maioria dos domínios, cada variável aleatória é diretamente influenciada por no max k outras vars (nós pais).
- Qtde necessária de números para a TPC de cada nó:  $2^k$ .
- Para a rede completa (n nós):  $n2^k$ .

## Compactação e Ordenação de Nós

- Ex concreto: rede com 20 nós e no max 5 pais p/ cada nó:
  - redes de crença: 640 números.
  - tabela de distr de prob conj: ordem de  $10^6$  números.
- Número de links extra na rede = maior precisão, mas pode não se justificar devido ao aumento do tamanho das tabelas.
- Regra geral: adicionar à rede primeiro os nós causadores de algum efeito e depois seus efeitos.

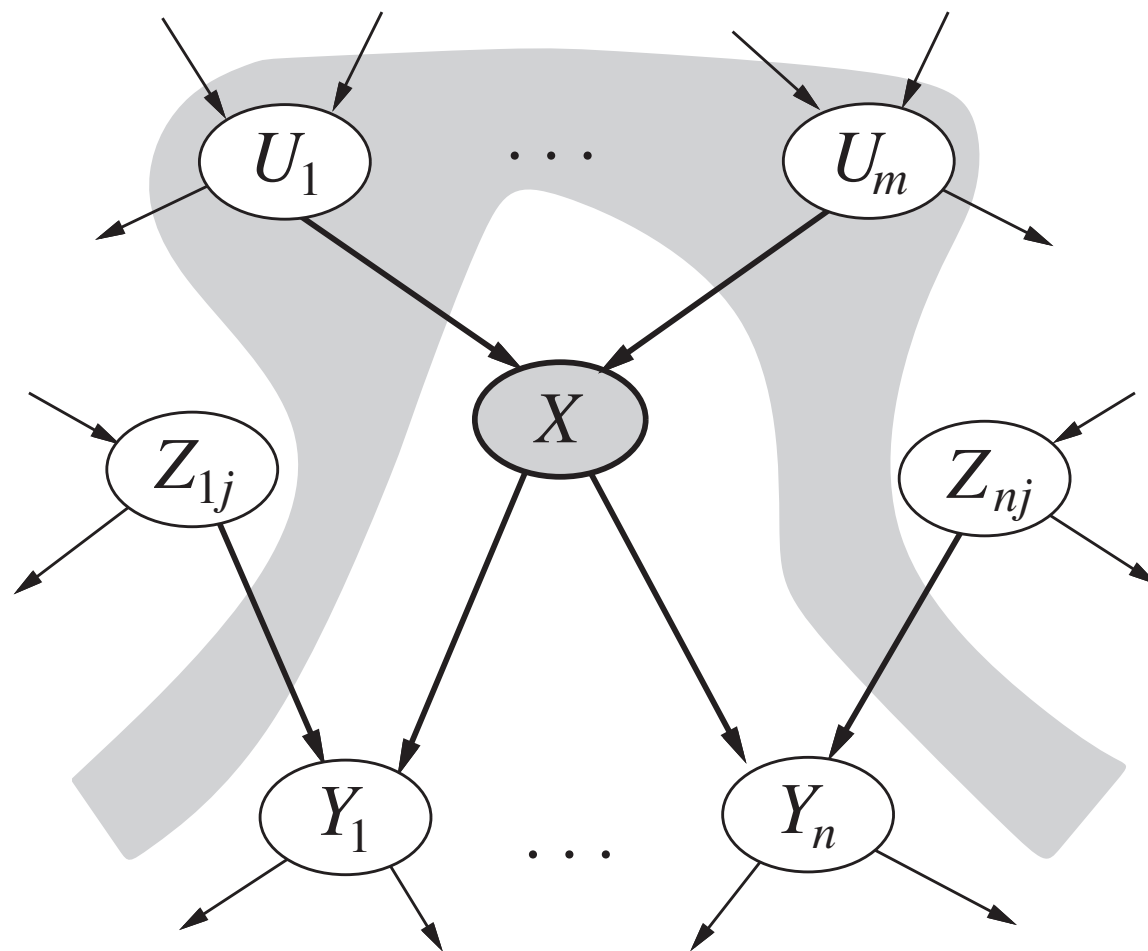
## Representação de TPCs

- Problema: escolher as probs condicionais das TPCs.
- Relação entre pais e filhos pode se encaixar numa distribuição canônica. Neste caso, probs podem ser especificadas através de nomes e talvez parms adicionais.
- Ex mais simples: nós determinísticos, probs são iguais as probs dos pais.
- Nós não determinísticos: relação *ruidosa* OU.
- Representação das probs:
  - Se todos os pais F, nó de saída F, com 100% de certeza.
  - Se apenas 1 dos pais é V, nó de saída é F com prob = parâmetro ruidoso daquele nó pai que é V.
- Ex:  $P(\text{Febre} \mid \text{Resf}) = 0.4$ ,  $P(\text{Febre} \mid \text{Gripe}) = 0.8$  e  $P(\text{Febre} \mid \text{Mal}) = 0.9$

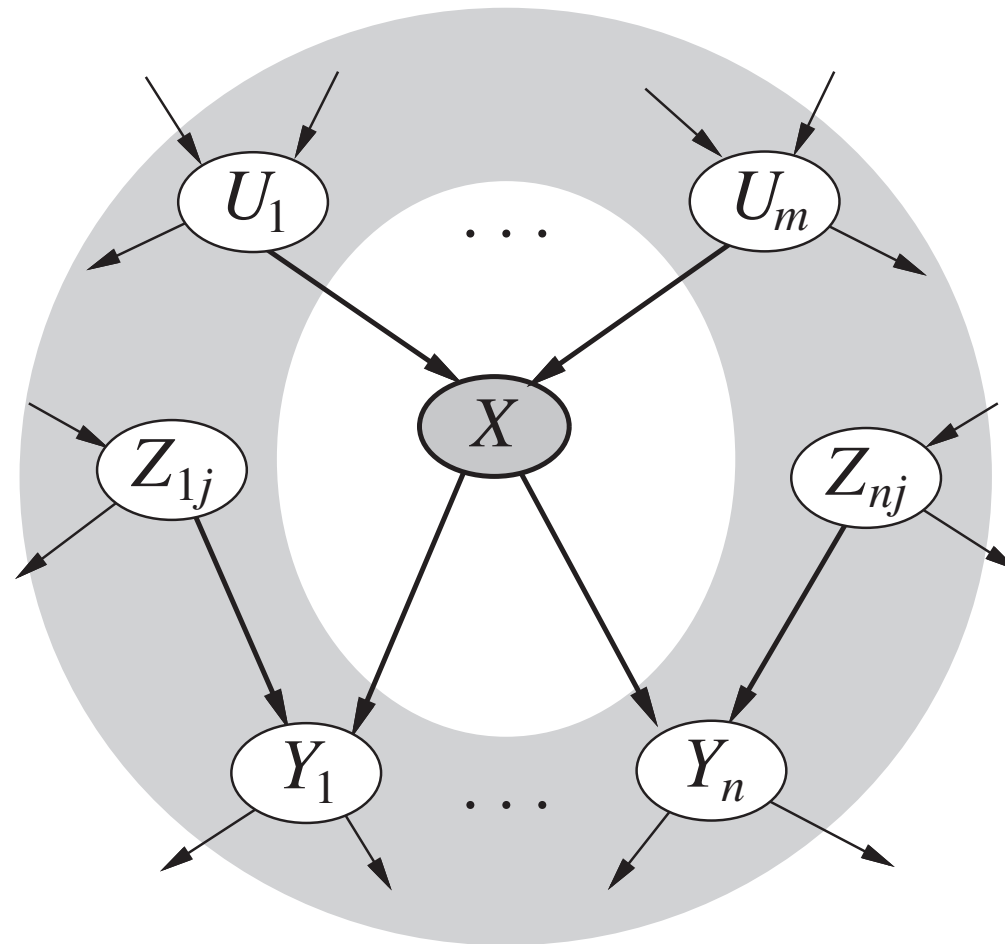
Resf	Gripe	Mal	P(Febre)	P( $\neg$ Febre)
F	F	F	0.0	1.0
F	F	V	0.9	0.1
F	V	F	0.8	0.2
F	V	V	0.98	$0.02 = 0.2 \times 0.1$
V	F	F	0.4	0.6
V	F	V	0.94	$0.06 = 0.6 \times 0.1$
V	V	F	0.88	$0.12 = 0.6 \times 0.2$
V	V	V	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

## Relações de Independência Condicional

- Necessidade: saber se relações de independência condicional mais gerais (não somente de pais p/ filhos) existem para poder obter mecanismos de inferência capazes de responder consultas do tipo: “Existe algum conj de nós X independente de outro conj Y, dado conj de evidências E?”



## Relações de Independência Condicional: Markov Blanket



## Inferência em Redes de Crenças

- Objetivo principal: computar distribuição de probabilidade posterior para um conj de variáveis de consulta, dados valores exatos para variáveis de evidência:  $P(Cons | Evidencia)$ .
- A princípio, qq nó pode servir como consulta ou evidência.
- Duas funções: BELIEF\_NET\_TELL p/ incluir novas evidências na rede e BELIEF\_NET\_ASK p/ computar novas probs p/ as variáveis de consulta.

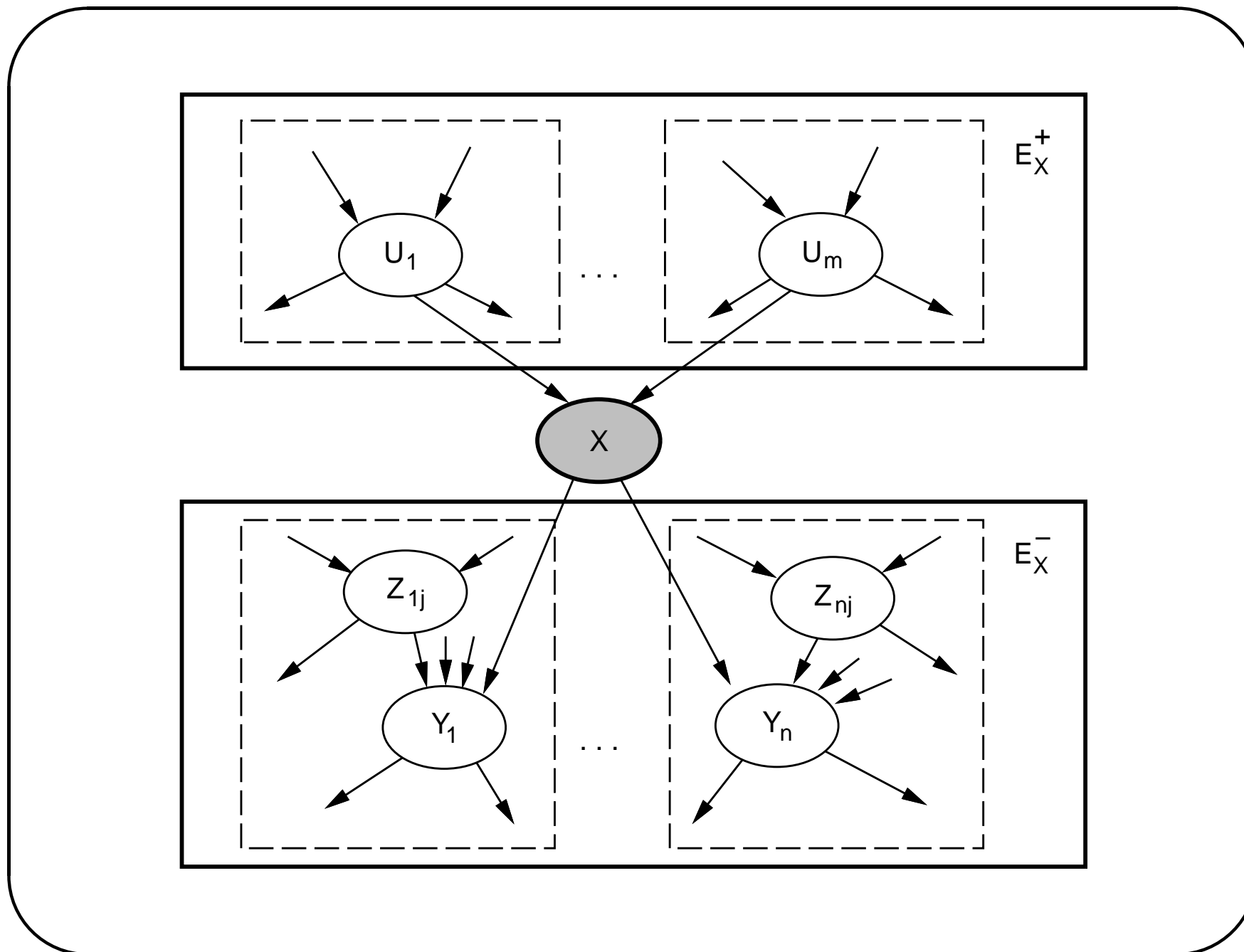


## Inferência em Redes de Crenças

- **Diagnósticas:**  $P(\text{Assalto} \mid \text{JohnTel})$  (efeitos p/ as causas).
- **Causais:**  $P(\text{JohnTel} \mid \text{Assalto})$  (causas p/ os efeitos).
- **Inter-causais:**  $P(\text{Assalto} \mid \text{Alarme} \wedge \text{Terr})$ .
- **Mistas:** combinação de 1 ou mais dos casos acima.
- Redes de crenças ainda podem ser usadas para:
  - tomar decisões baseadas em probs na rede e em funções de utilidade do agente.
  - decidir quais variáveis de evidência observar para obter info mais útil.
  - fazer “análise de sensibilidade” para entender aspectos do modelo que têm  $>$  impacto nas probs das vars de consulta.
  - Explicar os resultados da inferência probabilística para o usuário.

## Inferência em Redes de Crenças

- Algoritmo BELIEF\_NET\_ASK análogo ao backward-chain, faz inferências a partir das variáveis de consulta até encontrar alguma evidência.
- Algoritmo funciona somente para redes “singly connected”, onde há no máximo um ramo não dirigido entre quaisquer dois nós da rede: **poli-árvores**.
- Algoritmos para redes mais gerais usam algoritmos poli-árvore como sub-rotinas.



## Inferência em Redes de Crenças

- Nó  $X$  tem pais  $U$  e filhos  $Y$ .
- “singly connected” significa que todos os blocos são disjuntos e não têm links.
- $X$  é a variável de consulta.
- Objetivo: computar  $P(X | E)$ .
- Conjunto de **suporte causal**: variáveis de evidência “acima” de  $X$  que estão conectadas a  $X$  através dos pais.
- Conjunto de **suporte evidencial**: variáveis de evidência “abaixo” de  $X$  que estão conectadas a  $X$  através dos filhos.
- $E_{U_i|X}$ : evidências conectadas com todos os nós  $U_i$ , *exceto* via o ramo que passa por  $X$ .

## Inferência em Redes de Crenças

- Estratégia geral:
  - Representar  $\mathbf{P}(X | E)$  em termos de contribuições de  $E_X^+$  e  $E_X^-$ .
  - Computar a contribuição de  $E_X^+$  através do seu efeito nos pais de X. Obs: computar os efeitos dos pais de X em X pode ser feito de forma recursiva.
  - Computar a contribuição de  $E_X^-$  através do seu efeito nos pais de X. Obs: computar os efeitos dos pais de X em X pode ser feito de forma recursiva.
- Método: aplicar Bayes, outros métodos padrão de prob, simplificações (independência condicional).

## Inferência em Redes de Crenças

- $\mathbf{P}(X | E) = \mathbf{P}(X | E_X^-, E_X^+) = \frac{\mathbf{P}(E_X^- | X, E_X^+) \mathbf{P}(X | E_X^+)}{\mathbf{P}(E_X^- | E_X^+)}$
- Como X d-separa  $E_X^+$  de  $E_X^-$  na rede, podemos usar indep cond p/ simplificar primeiro termo do numerador. Tb podemos usar  $\frac{1}{\mathbf{P}(E_X^- | E_X^+)}$  como cte de normalização:  

$$\mathbf{P}(X | E) = \alpha \mathbf{P}(E_X^- | X) \mathbf{P}(X | E_X^+)$$
- $\mathbf{P}(X | E_X^+) = \sum_u \mathbf{P}(X | \mathbf{u}, E_X^+) P(\mathbf{u} | E_X^+)$
- $\mathbf{P}(X | E_X^+) = \sum_u \mathbf{P}(X | \mathbf{u}) \prod_i \mathbf{P}(u_i | E_X^+)$
- $\mathbf{P}(X | E_X^+) = \sum_u \mathbf{P}(X | \mathbf{u}) \prod_i \mathbf{P}(u_i | E_{U_i|X})$

## **Abordagens para aprendizagem da estrutura das redes**

- Restrições (constraint-based): tenta encontrar a melhor rede que satisfaz as dependências e independências entre as variáveis (através dos dados)
- Score: tenta encontrar a rede que tem melhor “score” (BIC, MDL, AIC etc)
- Híbrida: integra as abordagens constraint-based e score-based