Redes Bayesianas

- São diagramas que organizam o conhecimento numa dada área através de um mapeamento entre *causas* e *efeitos*
- Os sistemas baseados em redes Bayesianas são capazes de gerar automaticamente predições ou decisões mesmo na situação de inexistência de algumas peças de informação
- Marcos importantes:
  - » **1763 - Rev. Thomas Bayes**
  - » **Anos 80 - Judea Pearl**
  - » **1980 - investigadores dinamarqueses - métodos eficientes para abordar a ambiguidade ou falta de informação**
  - » **Microsoft: em 1993 contratam Eric Horvitz, David Heckerman e Jack Breese**
    - – **exploram esta abordagem nos seus produtos**
    - – **Microsoft Office usa esta tecnologia para fornecer ajuda ao utilizador baseado na historia passada, na forma como o rato se movimenta e na actividade em curso.**
      - **Se as acções mostram que o utilizador está distraído, é provável que necessite de ajuda. Se está a trabalhar num gráfico, a hipótese plausível é de que necessite de ajuda na formatação**
    - – **Investigação em técnicas de aprendizagem ou actualização das redes de uma forma automática baseada na experiência passada.**

- **O nome de redes Bayesianas deriva da utilização da fórmula matemática para o cálculo de probabilidades estabelecida pelo Rev. Thomas Bayes (1763)**

Abordagens Bayesianas ao Problema de Diagnóstico

- Nos anos 60 foram experimentadas técnicas de teoria da decisão e técnicas Bayesianas em problemas de diagnóstico
- Problemas:
  - » **Dificuldades na representação**
    - – **demasiadas probabilidades**
    - – **probabilidades não intuitivas**
  - » **Dificuldades na inferência**

  o raciocínio Bayesiano foi então considerado como um ideal não realizável

  - » **Soluções propostas:**
    - – **abordagens heurísticas**
    - – **factores de certeza**
    - – **lógica difusa**
    - – **teoria de Dempster& Shaffer**

- As Redes Bayesianas e os diagramas de influências renovaram o interesse nos métodos Bayesianos

Incerteza e Decisões Racionais

- **Teoria da probabilidade:**
  - » **proporciona uma medida ou grau de crença (credito) no conhecimento**
  - » **é uma forma de sumarizar a incerteza**

  - » **Exemplo: podemos não saber *a priori* qual o problema de um dado paciente, mas acreditamos existir uma probabilidade de 80% de ele sofrer de uma cárie dentária quando ele se queixa de dor de dentes**

(degree of belief) ***(degree of truth)***

grau de crença $\neq$ grau de verdade

(80% dos doentes analisados até ao momento com dores de dentes exibiam cárie dentária)

Incerteza e Decisões Racionais

Probabilidade — evidência

Todas as asserções probabilísticas devem traduzir a evidência com respeito ao que se conseguiu apurar

- **Antes de se obter alguma evidência falamos de *probabilidade a priori* ou *probabilidade não-condicionada***
- **Depois de obtida evidência falamos de *probabilidade a posteriori* ou *probabilidade condicionada***

- **Para se tomar uma decisão deve-se ter em conta as preferências (custos) sobre as possíveis situações**
  - » exemplos: falso alarme, falha de detecção, acerto, …

  ***Teoria da utilidade* - representa e manipula preferências**

- **As preferências, expressas por funções de utilidade ou custos, são combinadas com as probabilidades na teoria geral denominada *Teoria da Decisão***

Teoria da decisão = teoria da probabilidade
+
teoria da utilidade

Teoria da Decisão Bayesiana

- **Ideia fundamental:**
  - » **A decisão racional é a que escolhe a acção (decisão) que proporciona a maior utilidade esperada, calculada a média sobre todas as possíveis consequências dessa acção**

$$\Re = E\left[c(\omega,\hat{\omega})\right]$$
$$= E\left[\sum_i c(\omega_i,\hat{\omega})\, p(\omega_i \mid x)\, p(x)\right]$$

Distribuições de Probabilidade e Independência condicional

- **A distribuição conjunta para as variáveis x_1, …, x_n pode ser factorizada:**

$$p(x_1,\ldots,x_n \mid \xi) = p(x_n \mid x_1,\ldots,x_{n-1},\xi)\, p(x_1,\ldots,x_{n-1},\xi)$$

$$= p(x_1 \mid \xi)\, p(x_2 \mid x_1\xi)\, p(x_3 \mid x_1, x_2, \xi)\ldots$$

$$\ldots p(x_n \mid x_1,\ldots,x_{n-1},\xi)$$

$$= \prod_{\mathrm{i=1}}^{\mathrm{n}} p(x_i \mid x_1,\ldots,x_{i-1},\xi)$$

- **Dadas certas relações de independência condicionais**
  - »**independência condicional de x e y dado z:**
    - **p(x|y,z)=p(x|z)**

  a distribuição acima pode ser simplificada

$$p(x_1,\ldots,x_n \mid \xi) = \prod_{\mathrm{i=1}}^{\mathrm{n}} p(x_i \mid \pi_i, \xi)$$

em que $\pi_i \subseteq \{x_1,\ldots,x_{i-1}\}$ **é o conjunto de variáveis que torna** x_i **e** $\{x_1,\ldots,x_{i-1}\}$ **condicionalmente independentes**

para cada x_i, $\pi_i \subseteq \{x_1,\ldots,x_{i-1}\}$:

$$p(x_i \mid x_1,\ldots,x_{i-1} \mid \xi) = p(x_i \mid \pi_i, \xi)$$

- **As redes Bayesianas são estruturas que representam a dependência entre variáveis, dando uma especificação concisa da distribuição conjunta**

- **A *estrutura* de uma rede Bayesiana codifica as relações de independência condicionada**

- **O *conjunto de probabilidades* de uma rede Bayesiana é a colecção de distribuições locais**

$$p(x_i \mid \pi_i, \xi)$$

para cada nó do domínio

- **Uma rede Bayesiana é definida pela sua estrutura e modelo probabilístico, determinando de forma unívoca a distribuição conjunta para as variáveis que descreve**

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid pais(x_i))$$

$$= \prod_{i=1}^{n} p(x_i \mid \pi_i)$$

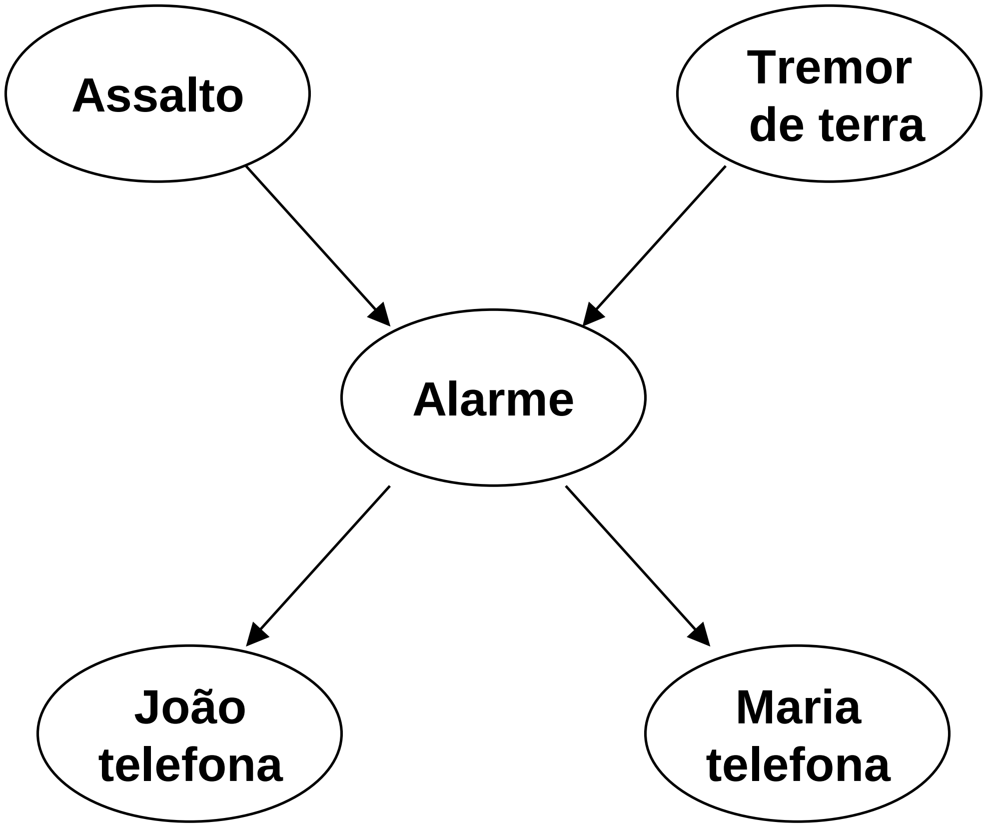
Representação do Conhecimento usando Redes Bayesianas

- **Rede Bayesiana**
  - » ***belief network***
  - » ***Bayesian network***
  - » ***probabilistic network***
  - » ***causal network***
  - » ***knowledge map***

- **Definição: É um grafo direccionado acíclico com as seguintes características**
  - » **Os nós correspondem a variáveis aleatórias**
  - » **Uma ligação direccionada ou arco com seta liga pares de variáveis (nós). O significado intuitivo de um arco dirigido do nó X para o nó Y é que *X tem uma influência directa sobre Y***
  - » **Cada nó tem associados os estados da variável que representa e uma tabela de probabilidades condicionadas que quantifica os efeitos que os pais exercem sobre um nó (probabilidade do nó estar num estado específico dado os estados dos seus pais)**
  - » **O grafo não possui ciclos direccionados**

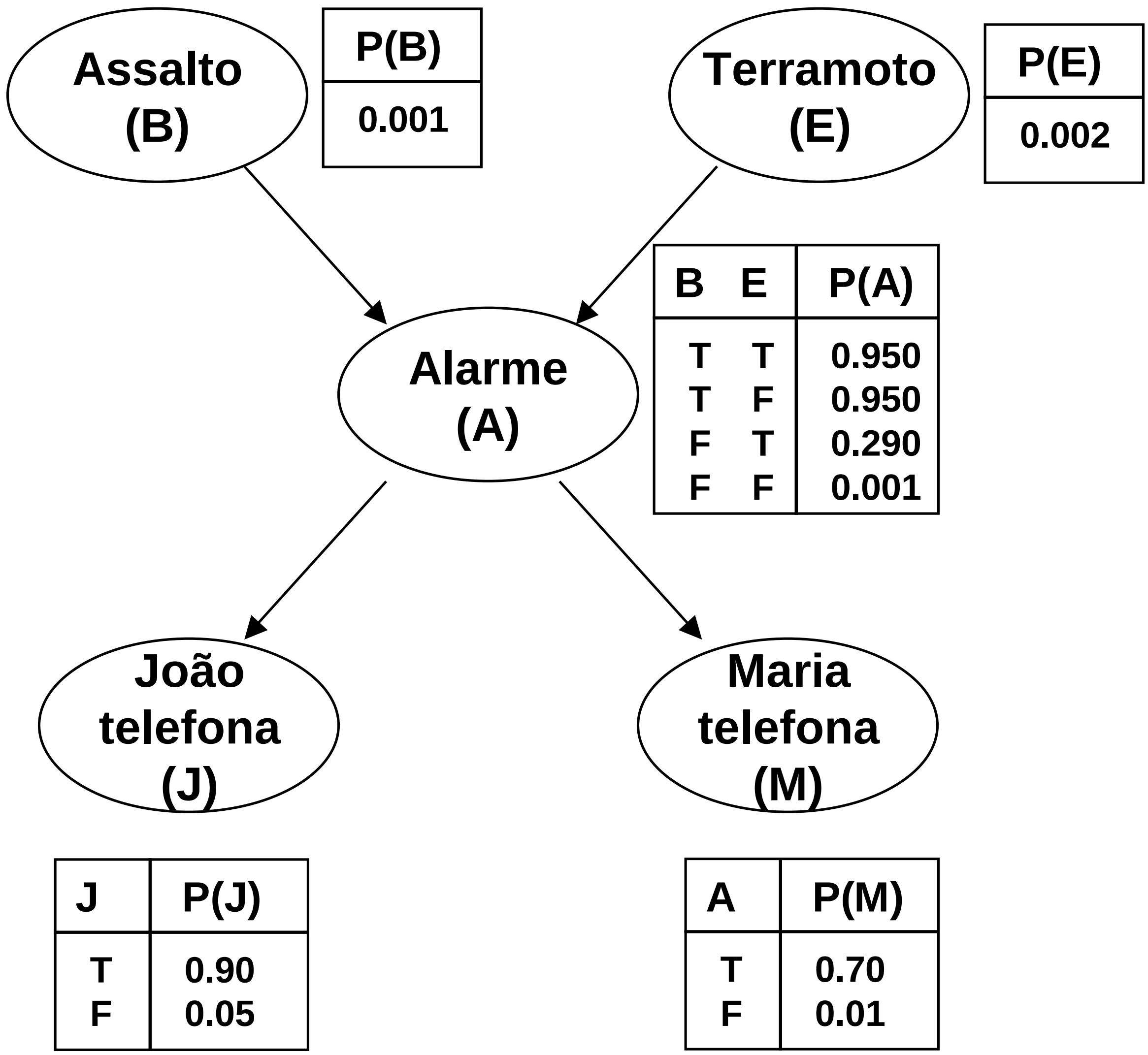
- **A topologia da rede pode ser vista como uma base de conhecimento abstracta, representando a estrutura dos processos causais no domínio.**

- **Exemplo - Judea Pearl - LA**
- **Situação:**
  - » **Sistema de alarme instalado em casa**
    - – **é activado com grande fiabilidade pela intrusão de estranhos - assalto**
    - – **também responde quando existem pequenos tremores de terra**
  - » **Dois vizinhos que prometem telefonar para o emprego quando ouvirem o alarme**
    - – **Maria: costuma ouvir música muito alta e por vezes não ouve o alarme**
    - – **João: telefona sempre que ouve o alarme mas por vezes confunde o alarme com o toque do telefone**
- **Questão:**
  - » **Dada a evidência de que alguém telefonou ou não telefonou, gostaríamos de estimar a probabilidade de haver um assalto**



Uma vez especificada a topologia da rede, é necessário especificar a tabela de probabilidades condicionadas para cada nó

		P(Alarme|Ass.,Terr.)	
Assalto	**Terramoto**	***Verdade***	***Falso***
Verdade	Verdade	0.950	0.050
Verdade	falso	0.950	0.050
Falso	Verdade	0.290	0.710
Falso	Falso	0.001	0.999



P(B)
0.001

P(E)
0.002

B	E	P(A)
T	T	0.950
T	F	0.950
F	T	0.290
F	F	0.001

J	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

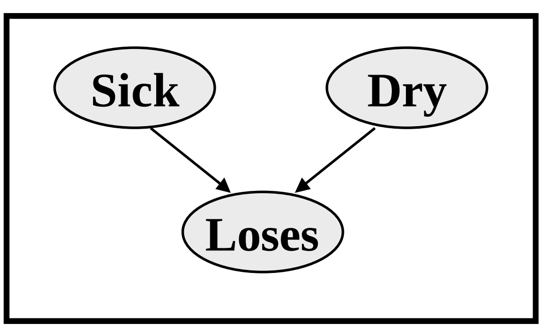
Distribuição conjunta de probabilidade:

$$p(x_1,\ldots,x_n) = \prod_{\mathrm{i=1}}^{\mathrm{n}} p(x_i \mid pais(x_i))$$

$$\begin{aligned} &P(J, M, A, \neg B, \neg E) = \\ &= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062 \end{aligned}$$

Exemplo da “Apple Tree” - HUGIN

- **Domínio: pomar de macieiras do senhor Jack Fletcher.**
  - » **Um dia o senhor Jack descobre que uma das macieiras está a perder folhas e deseja saber porquê.**
  - » **Sabe que se as árvores tiverem pouca água (seca…) é muito comum caírem as folhas; no entanto o mesmo acontece quando as árvores estão doentes**
- **Modelação através de RB:**
  - » **nós:**
    - – ***Sick*** **- estados: *sick* e *not sick***
    - – ***Dry*** **- estados: *dry* e *not dry***
    - – ***Loses*** **- estados: *yes* e *no***
  - » ***Rede:*** **dependência causal de *sick* para *loses* e de *dry* para *loses* é representada pelos arcos**



- **Pergunta: estará a árvore doente dada a evidência de que está a perder folhas?**
  - » **Conhecido: P(sick) ; P(dry) ; P(loses|sick,dry)**
  - » **Evidência: *loses=yes***
  - » **Questão: P(sick|loses=yes)**

$$P(sick \mid loses = yes) = \frac{P(loses = yes \mid sick)P(sick)}{P(loses = yes)}$$

$$P(loses = yes) \quad = \sum_{\substack{dry \\ sick}} P(loses = yes \mid dry, sick)P(dry, sick)$$

$$= \sum_{\substack{dry \\ sick}} P(loses = yes \mid dry, sick)P(dry)P(sick)$$

$$= P(loses = yes \mid dry = y, sick = y)P(dry = y)P(sick = y) +$$

$$+ P(loses = yes \mid dry = y, sick = n)P(dry = y)P(sick = n) +$$

$$+ P(loses = yes \mid dry = n, sick = n)P(dry = n)P(sick = n) +$$

$$+ P(loses = yes \mid dry = n, sick = y)P(dry = n)P(sick = y)$$

$$= 0.95 * 0.1 * 0.1 + 0.85 * 0.1 * 0.9 + 0.02 * 0.9 * 0.9 + 0.9 * 0.9 * 0.1$$

$$= 0.1832$$

$$
\begin{aligned}
&P(loses = yes \mid sick = y) \\
&\quad = \sum_{dry} P(loses = yes \mid dry, sick = y) P(dry \mid sick) \\
&\quad = \sum_{dry} P(loses = yes \mid dry, sick = y) P(dry) \\
&\quad = P(loses = y \mid sick = y, dry = y) P(dry = y) + \\
&\quad \quad + P(loses = y \mid sick = y, dry = n) P(dry = n) \\
&\quad = 0.95 * 0.1 + 0.9 * 0.9 = 0.9050
\end{aligned}
$$

$$
\begin{aligned}
&P(loses = yes \mid sick = n) \\
&\quad = \sum_{dry} P(loses = yes \mid dry, sick = y) P(dry \mid sick) \\
&\quad = \sum_{dry} P(loses = yes \mid dry, sick = n) P(dry) \\
&\quad = P(loses = y \mid sick = n, dry = y) P(dry = y) + \\
&\quad \quad + P(loses = y \mid sick = n, dry = n) P(dry = n) \\
&\quad = 0.85 * 0.1 + 0.02 * 0.9 = 0.1030
\end{aligned}
$$

$$
\begin{aligned}
P(sick = y) \mid loses = y) \quad &= \frac{P(lose = y \mid sick = y) P(sick = y)}{P(loses = y)} \\
&= \frac{0.9050 * 0.1}{0.1832} = 0.4940
\end{aligned}
$$

$$
\begin{aligned}
P(sick = n) \mid loses = y) \quad &= \frac{P(lose = y \mid sick = n) P(sick = n)}{P(loses = y)} \\
&= \frac{0.1030 * 0.9}{0.1832} = 0.5060 \\
&= 1 - P(sick = y \mid loses = y)
\end{aligned}
$$

Construção de uma Rede Bayesiana

- **Dada a distribuição conjunta e a definição de probabilidade condicionada, a primeira pode ser descrita como**

$$p(x_1,\dots,x_n) = p(x_1)\,p(x_2 \mid x_1)\,p(x_3 \mid x_1, x_2)\dots\, p(x_n \mid x_1,\dots,x_{n-1})$$

$$= \prod_{i=1}^{n} p(x_i \mid x_1,\dots,x_{i-1})$$

- **Comparando com a expressão para as redes Bayesianas**

$$p(x_1,\dots,x_n) = \prod_{i=1}^{n} p(x_i \mid pais(x_i))$$

a equivalência entre as expressões significa que

$$p(x_i \mid x_1,\dots,x_{i-1}) = p(x_i \mid pais(x_i))$$

- **A relação anterior diz que a rede Bayesiana é a representação correcta para o domínio quando cada nó é condicionalmente independente dos seus antecessores na ordenação dos nós dados os seus pais.**

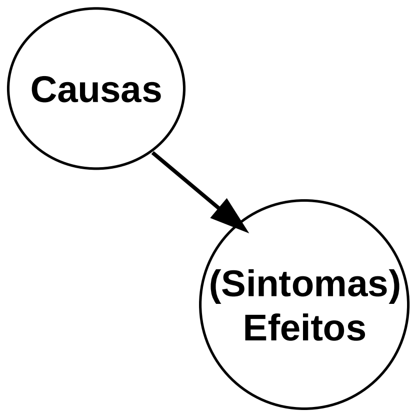
- **Assim, para construír uma rede Bayesiana com a estrutura correcta para o domínio é necessário escolher os pais para cada nó de forma a que esta propriedade se verifique**
  - » **os pais de um nó x_i devem incluir todos os nós $x_1,\ldots, x_{i-1}$ que influênciam directamente x_i**
- ***São as relações de independência condicionadas entre as variáveis que devem guiar a construção da topologia da rede***

- **Método:**
  - » **Escolher o conjunto de variáveis relevantes x_i que descrevem o domínio**
  - » **Escolher uma ordenação para as variáveis**
  - » **Enquanto existirem variáveis:**
    - – **seleccionar uma variável x_i e adicionar um nó à rede para esta**
    - – **definir os *pais*(x_i) como o conjunto mínimo de nós já existentes na rede para os quais a propriedade de independência condicionada se verifique**
    - – **definir a tabela de probabilidade condicionada para x_i**

- **Uma vez que cada nó apenas se liga a nós definidos anteriormente, este método de construção garante que a rede é acíclica**

Ordenação dos Nós

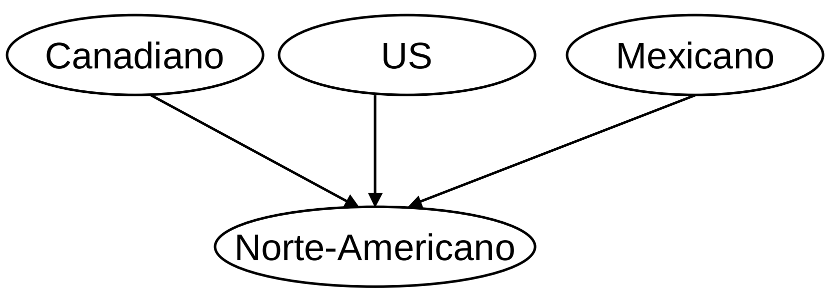
- **A ordem correcta para adicionar nós consiste em começar por adicionar as *causas* - nós raíz da rede - e de seguida as variáveis que estas influenciam, sucessivamente até se atingirem as folhas da rede (variáveis que não possuem uma relação causal com nenhuma outra)**



- **Quando se tenta construir o modelo com ligações dos sintomas para as causas existe a necessidade de especificar dependências adicionais (redes mais complexas) que requerem por vezes a definição de probabilidades que são difíceis de obter ou não naturais**
- **A escolha da ordenação de *causas* para *sintomas* conduz em geral a redes mais compactas e de mais fácil definição em termos de probabilidades**
- **No domínio da medicina os médicos preferem explicitar os dados estatísticos para *regras causais* do que para *regras de diagnóstico***

Tabelas de Probabilidades

- **O preenchimento das tabelas de probabilidades condicionadas é muitas vezes simples (desde que a relação entre os pais e o nó filho não seja arbitrária)**
- **Geralmente as relações entre nós pais e nós filhos caem em categorias de *distribuições canónicas* (que obedecem a um padrão), sendo necessário apenas identificar qual o padrão e introduzir alguns parâmetros**
  - » ***Nós determinísticos* - têm o seu valor especificado a partir dos valores dos pais, sem incerteza (p=1, ou p=0)**
    - – **relação lógica: Exexmplo: disjunção**



C U M	P(NA)
T T T	1
T T F	1
...	...
F F F	0

– **Relação numérica. Exemplo:**
  - **Pais: preço de um dado modelo de carro em vários stands**
  - **Filho: Preço negociado; o valor é o mínimo dos valores dos pais, sem incerteza**

» ***Noisy-OR*** **- generalização do OR com incerteza.**

– **cada causa tem uma probabilidade independente de provocar o efeito**

– **todas as possíveis causas estão listadas (*leak-node*: causas miscelaneas)**

– **o que iniba a conclusão é independente de outras inibições, sendo as inibições modeladas por *noise parameters*; as inibições combinam-se então por multiplicação:**
  - **se nenhum pai é *verdade* então o filho é *falso* com 100% de certeza**
  - **se apenas um dos pais é *verdade* o filho é *falso* com probabilidade igual ao *noise-parameter* para esse nó**
  - **em geral a probabilidade do nó filho ser *falso* é dado pelo produto dos *noise-parameter*s para cada nó pai que é *verdade***

– **Exemplo:**
  - **lógica proposicional:**
    - » ***Febre* é verdade sse *Constipação, Gripe* ou *Malaria***
  - **noisy-or:**
    - » **P(*Febre* | *Constipação*)=.4; P(*Febre*|)=.8**
    - » **P(*Febre* |*Malaria)*=.9**
    - » ***noise parameters:* .6 ; .2 ; .1**

Const	*Gripe*	*Malaria*	P(*Febre*)	P(¬*Febre*)
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=0.2x0.1
T	F	F	0.4	0.6
T	F	T	0.94	0.06=0.6x0.1
T	T	F	0.88	0.12=0.6x0.2
T	T	T	0.988	0.012=0.6x0.2x0.1

Compacticidade

- **As redes Baysesianas são representações completas e não redundantes de um domínio, sendo em geral mais compactas do que a explicitação directa da distribuição conjunta**
  - » **Isto deve-se às características de estruturação local da rede, expressando as características de independência condicionadas entre variáveis**
- **Estruturação local da rede:**
  - » **cada sub-componente (nó) interage directamente únicamemente com um número reduzido de outros componentes**
  - » **Suponhamos que num dado domínio cada variável é directamente influenciada por um máximo de *k* variáveis; se as variáveis forem booleanas:**
    - – **é necessário 2^k números para descrever a tabela de probabilidade condicionada para cada nó**
    - – **a rede completa é descrita por $n2^k$ números (n = nº de nós)**
    - – **Descrição da tabela da distribuição conjunta: 2^n números**
  - » **Uma redução na representação significa uma boa captação da estrutura do domínio**

Inferência usando Redes Bayesianas

- A distribuição conjunta pode ser usada para responder a qualquer pergunta sobre o domínio
- As redes Bayesianas, como representação gráfica desta distribuição, podem também ser usadas para responder a qualquer questão
- Tipos de inferência:
  - » diagnóstico (dos efeitos para as causas)
    - – *P(Assalto|João telefona)=0.016*
  - » causais (das causas para os efeitos)
    - – *P(João telefona| Assalto)=0.86*
  - » intercausais (entre causas e um efeito comum)
    - – *P(Assalto|Alarme)=0.376*
    - – *P(Assalto|Alarme,Terramoto)=0.003*
  - » mistas (combinação das anteriores)

- Para além de responder a questões sobre probabildades de variáveis, a RB pode ainda ser usada para:
  - » Tomar decisões baseadas nas probabilidades na rede e em funções de custo / utilidade
  - » decidir que variáveis adicionais é necessário observar para se dispor de informação útil
  - » Realizar análise de sensibilidade
  - » Explicar os resultados da inferência probabilística

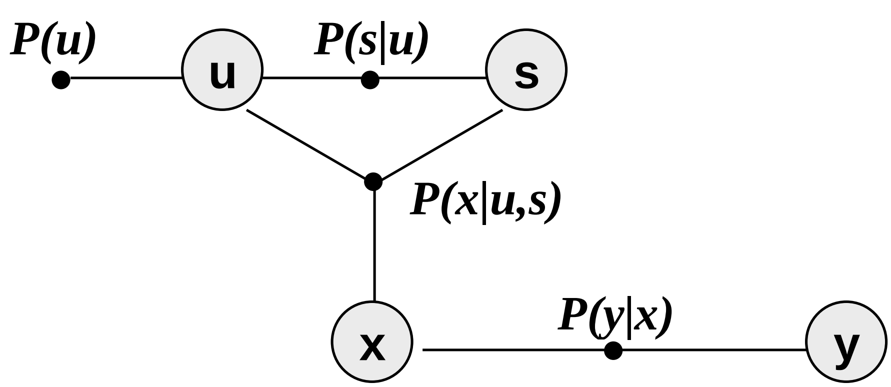
Inferência usando Redes Bayesianas

- **Inferência exacta usando propagação de probabilidades**
- **Algoritmo: *Generalized Forward-Backward***

  » ***Factor Graphs* (Grafos de factorização da distribuição conjunta)**

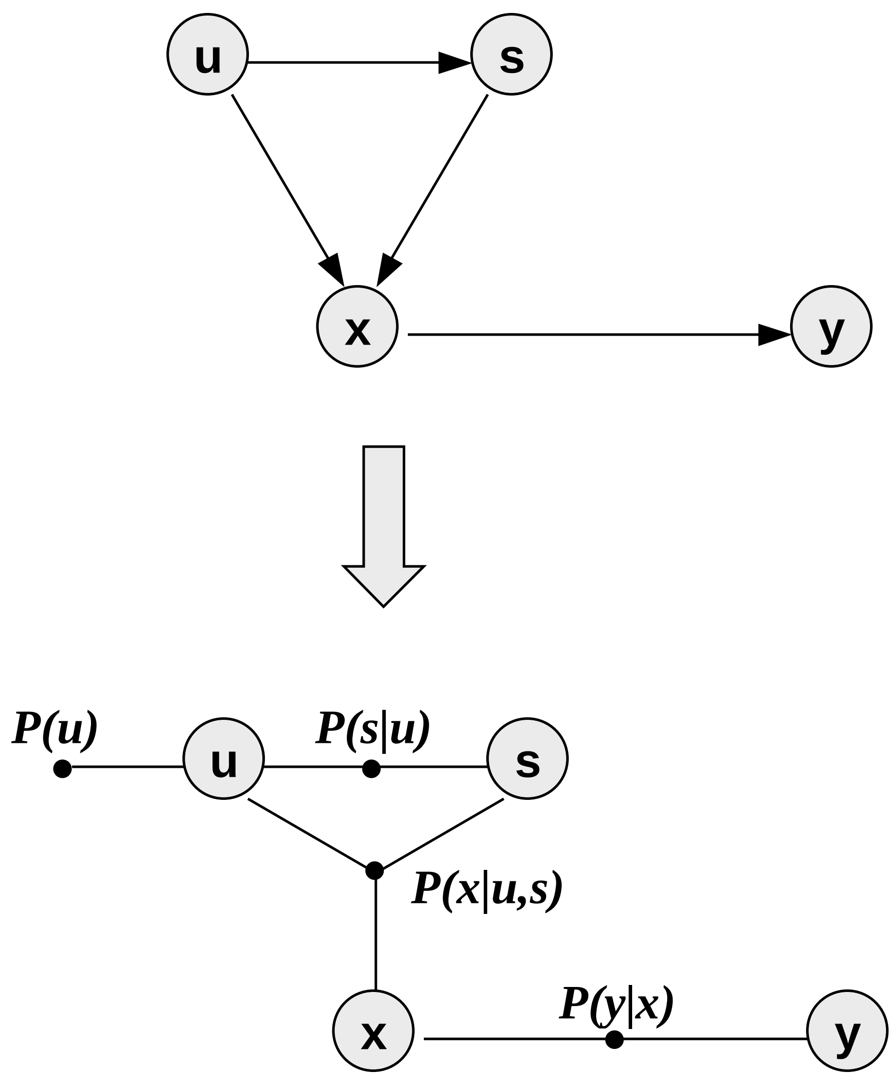
  – **indicam como a distribuição conjunta de muitas variáveis se pode factorizar como o produto de funções num conjunto mais reduzido de variáveis**

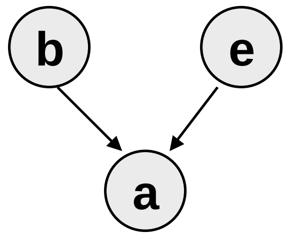
  – **Um grafo de factorização é um grafo com dois tipos de nós - correspondendo a variáveis ou a funções locais (distribuições condicionadas) - e em que os arcos só podem ligar nós de tipos diferentes. Cada função local está ligada às variáveis da qual depende.**



$$P(u,s,x,y) = P(u)P(s \mid u)P(x \mid u,s)P(y \mid x)$$

» **Uma rede Bayesiana pode ser escrita como um grafo de factorização por introdução de um nó de função por cada função de distribuição**

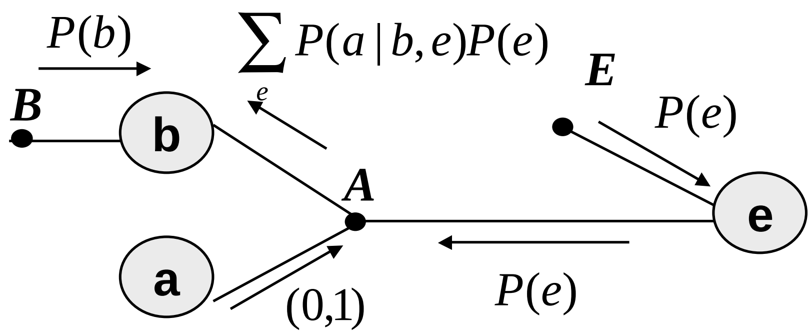




$$P(b\,|\,a)=\sum_e P(b,e\,|\,a)=\frac{\sum_e P(a\,|\,b,e)P(b,e)}{P(a)}=$$

$$=\frac{\sum_e P(a\,|\,b,e)P(e)P(b)}{P(a)}=\frac{\sum_e P(a\,|\,b,e)P(b)P(e)}{\sum_{b',e'} P(a\,|\,b',e')P(b')P(e')}$$

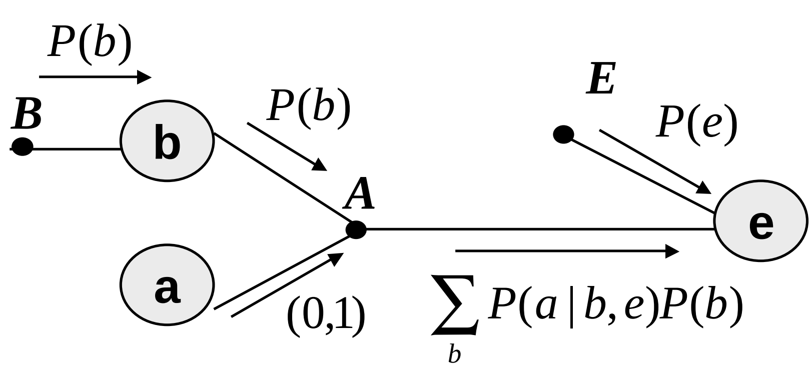
$$=\beta\sum_e P(a\,|\,b,e)P(e)\;P(b)$$



$$P(e\,|\,a)=\sum_b P(b,e\,|\,a)=\frac{\sum_b P(a\,|\,b,e)P(b,e)}{P(a)}=$$

$$=\frac{\sum_b P(a\,|\,b,e)P(b)P(e)}{P(a)}=\frac{\sum_b P(a\,|\,b,e)P(b)P(e)}{\sum_{b',e'} P(a\,|\,b',e')P(b')P(e')}$$

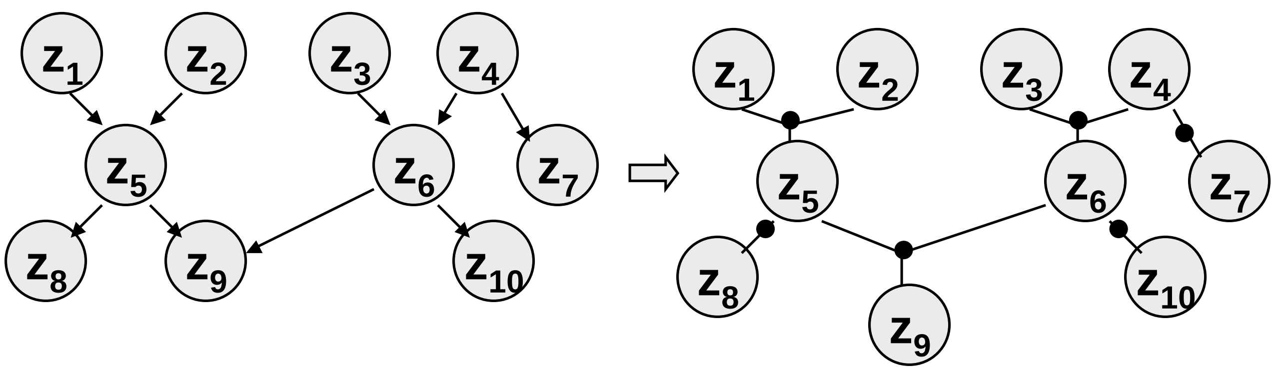
$$=\beta\sum_b P(a\,|\,b,e)P(b)\;P(e)$$



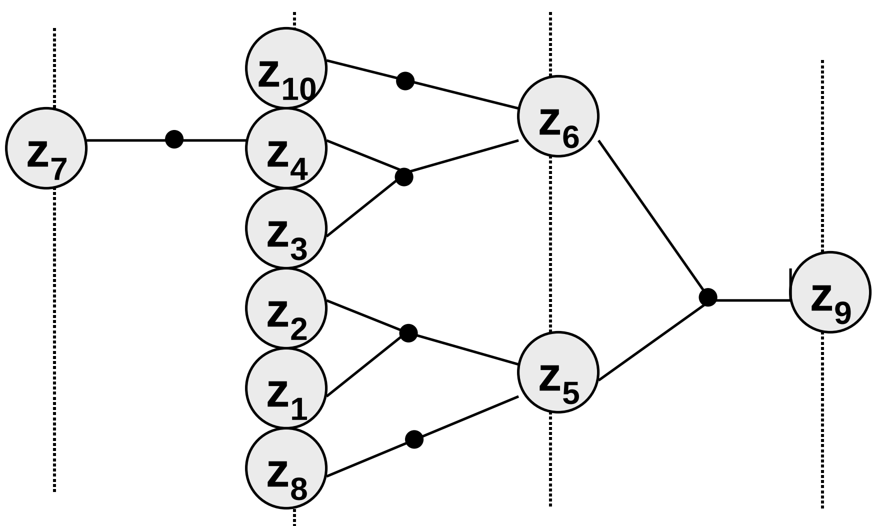
Algoritmo: *Generalized Forward-Backward*

- **Para redes do tipo *Singly connected* (não existe mais do que um trajecto a ligar quaisquer dois nós (em que as direcções dos arcos não são tomadas em conta)**

- **1º: Converter a rede Bayesiana (singly connected) num grafo de factorização**



- **2ª: Dispôr o grafo de factorização como uma árvore horizontal com uma raíz, arbitrariamente escolhida, colocada no extremo direito**



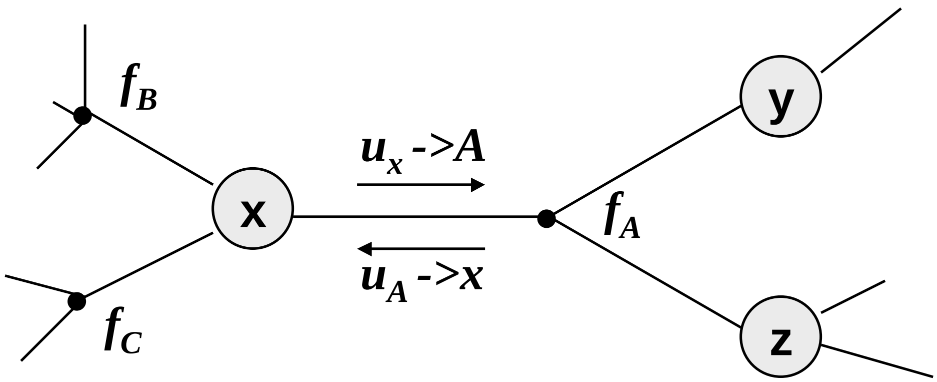
- **Começando no nível mais à esquerda da árvore, as mensagens são passadas em frente para a raíz, enquanto são multioplicadas e marginalizadas no processo. Cada nó armazena as suas mensagens para uso posterior.**
- **Depois as mensagens são passadas para trás, nível a nível, a partir da raíz até às folhas da árvore.**
- **Durante as duas fases as variáveis observadas são mantidas constantes.**
- **Uma vez completadas as duas fases, cada nó combina as mensagens que recebeu e armazenou para obter** $P(z_i|V)$
- **Durante cada uma das fase de *forward* e *backward* são passados dois tipos de mensagens:**
  - » **variável-para-função, consistindo de produtos de marginais locais**
  - » **função-para-variável, consistindo em somas de distribuições sobre variáveis circundantes**
- **Após completadas as fases acima a distribuição de x condicionada às observações V é dada por**

$$P(x\,|\,V) = \beta\mu_{A\to x}(x)\mu_{B\to x}(x)\mu_{C\to x}(x)$$

- **em que b é calculada para garantir a normalização**

$$\sum_x P(x\,|\,V) = 1$$

- **Ex.: Considere-se o fragmento do grafo:**



» **Mensagem variável-para-função $u_{x \to A}(x)$**

$$u_{x \to A}(x) = u_{B \to x}(x) * u_{C \to x}(x)$$

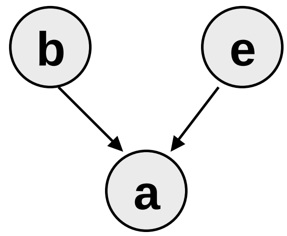
a menos que x seja observado, sendo neste caso

$$u_{x \to A}(x) = \delta(x, x')$$

» **Mensagem função -para-variável $u_{A \to x}(x)$**

$$\mu_{A\to x}(x) = \sum_y \sum_z f_A(x,y,z)\mu_{y\to A}(y)\mu_{z\to A}(z)$$

Exemplo



- e= *earthquake*
- *b=burglar*
- *a=alarm*

P(b=1)=0.1 ***P(e=1)=0.1***

P(a=1|b=0,e=0)=0.001

P(a=1|b=1,e=0)=0.368

P(a=1|b=0,e=1)=0.135

P(a=1|b=1,e=1)=0.607

- **Solução exacta usando directamente a regra de Bayes:**

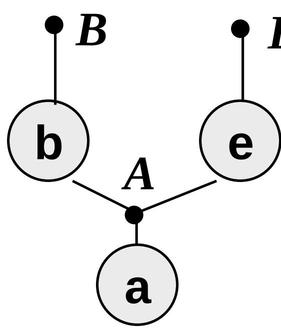
$$P(b,e\,|\,a)=\frac{P(a\,|\,b,e)P(b)P(e)}{\sum_{b',e'}P(a\,|\,b',e')P(b')P(e')}$$

$$\Rightarrow P(b=0,e=0\,|\,a=1)=\frac{0.001\times 0.9\times 0.9}{c}=0.016$$

$$c=0.001\times 0.9^2+0.368\times 0.1\times 0.9+$$
$$+0.15\times 0.1\times 0.9+0.607\times 0.1^2=0.05215$$

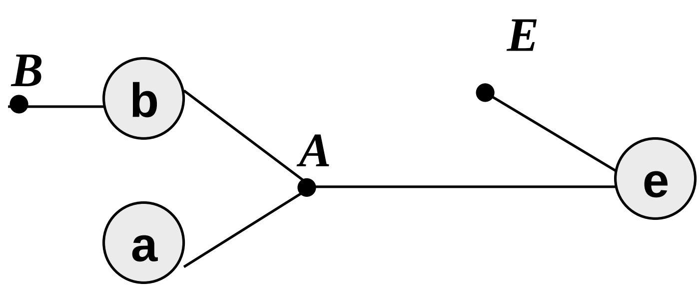
$$P(b=1,e=0\,|\,a=1)=0.635$$
$$P(b=0,e=1\,|\,a=1)=0.233$$
$$P(b=1,e=1\,|\,a=1)=0.116$$

- **Solução exacta usando o algoritmo *forward-backward:***
  - » **reescrita da rede como um grafo de factorização**



$$f_B(b) = P(b)$$
$$f_E(e) = P(e)$$
$$f_A(a,b,e) = P(a \mid b,e)$$

  - » **escolha arbitrária da raíz, por exemplo *e*, e organização em árvore horizontal**

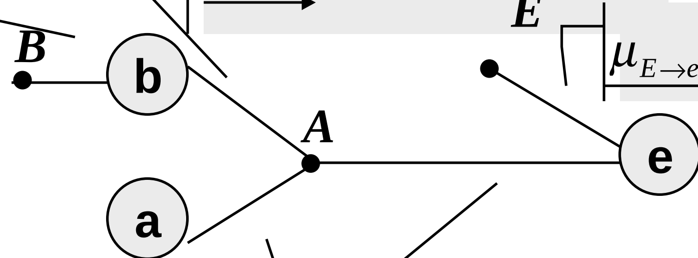


- **Envio de mensagens na direcção para a frente:**

$$\overrightarrow{\mu_{B\to b}} = P(b) = (0.9, 0.1)$$

$$\overrightarrow{\mu_{b\to A}} = \mu_{B\to b} = (0.9, 0.1)$$

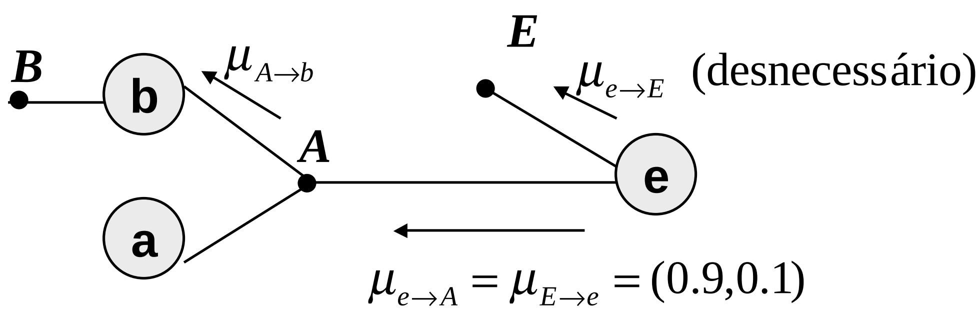
$$\overrightarrow{\mu_{E\to e}} = P(e) = (0.9, 0.1)$$



$$\overrightarrow{\mu_{a\to A}} = (0,1)$$

$$\mu_{A\to e} = \sum_{b,a} f_A(a,b,e)\mu_{b\to A}(b)\mu_{a\to A}(a)$$
$$= \sum_{b} P(a=1 \mid b,e)\mu_{b\to A}(b)$$
$$= (P(a=1 \mid b=0, e=0)\times P(b=0) +$$
$$P(a=1 \mid b=1, e=0)\times P(b=1),$$
$$P(a=1 \mid b=0, e=1)\times P(b=0) +$$
$$P(a=1 \mid b=1, e=1)\times P(b=1))$$
$$= (0.001\times 0.0 + 0.368\times 0.1,$$
$$0.135\times 0.9 + 0.697\times 0.1)$$
$$= (0.0377, 0.1822)$$

- **Envio de mensagens na direcção para trás:**



$$
\begin{aligned}
\mu_{A\to b}(b) &= \sum_{e,a} f_A(a,b,e)\mu_{e\to A}(e)\mu_{a\to A}(a) \\
&= \sum_{e} P(a=1\,|\,b,e)\mu_{e\to A}(e) \\
&= (P(a=1\,|\,b=0,e=0)\times P(e=0) + \\
&\quad P(a=1\,|\,b=0,e=1)\times P(e=1), \\
&\quad P(a=1\,|\,b=1,e=0)\times P(e=0) + \\
&\quad P(a=1\,|\,b=1,e=1)\times P(e=1) \\
&= (0.001\times 0.9 + 0.135\times 0.1, \\
&\quad 0.368\times 0.9 + 0.607\times 0.1) \\
&= (0.0144, 0.3919)
\end{aligned}
$$

- **Combinando os resultados:**

$$
\begin{aligned}
P(b\,|\,a=1) &= (P(b=0\,|\,a=1), P(b=1\,|\,a=1)) \\
&= \beta\times(\mu_{B\to b}(0)\times\mu_{A\to b}(0), \mu_{B\to b}(1)\times\mu_{A\to b}(1)) \\
&= (0.249, 0.751)
\end{aligned}
$$

$$P(e \mid a=1) = (P(e=0 \mid a=1), P(e=1 \mid a=1))$$
$$= \beta \times (\mu_{E\to b}(0) \times \mu_{A\to e}(0), \mu_{E\to b}(1) \times \mu_{A\to e}(1))$$
$$= (0.651, 0.349)$$

Estas distribuições são as mesmas que se obtêm através da marginalização do resultado obtido pela regra de Bayes:

$$P(b \mid a=1) = \sum_e P(b, e \mid a=1)$$
$$= (P(b=0, e=0 \mid a=1) + P(b=0, e=1 \mid a=1),$$
$$P(b=1, e=0 \mid a=1) + P(b=1, e=1 \mid a=1))$$
$$= (0.016 + 0.233, 0.635 + 0.116)$$
$$= (0.249, 0.751)$$

Propagação de Probabilidade Algoritmo *Sum-Product*

- **A forma regular de propagação das mensagens no algoritmo *generalized forward-backward* pode ser simplificado para obter um algoritmo mais geral de propagação de probabilidades. Desde que seja observado um conjunto mínimo de regras, as mensagens podem ser propagadas por qualquer ordem para obtenção das probabilidades condicionadas. Para além destas regras, as fórmulas de propagação são idênticas às do algoritmo *generalized forward-backward:***
  - » **Antes de começar a fase de propagação, inicialize-se a rede Bayesiana. Consiste no cálculo das probabilidades *a priori* para cada nó $P(z_i)$**
  - » **Criar mensagens como resposta a observações. Se a variável *y* é observada com o valor *y'*, então enviar uma mensagem para cada nó associado a *y* através de todos os arcos que lhe estão ligados usando**

$$\mu_{y \to A}(y) = \delta(y, y')$$

  - » **Propagar mensagens em resposta a outra mensagens. Se uma variável *y* recebe uma mensagem através de um dos arcos, então *y* deve enviar mensagens através de todos os outros arcos; o mesmo acontece se *y* corresponde a uma função**

» **As mensagens são absorvidas em nós ligados por um único arco**

» **Em qualquer instante do processo de propagação, o nó *y* pode calcular uma estimativa** $\hat{P}(y \mid V)$ **de** $P(y \mid V)$ **usando**

$$P(y \mid V) = \beta \mu_{A \to y}(y) \mu_{B \to y}(y) \mu_{C \to y}(y)$$

» **Se as regras acima forem observadas e se a propagação continuar até não haverem mais mensagens por propagar na rede, então as estimativas acima coincidem com as probabilidades exactas**

$$\hat{P}(y \mid V) = P(y \mid V)$$

Redes Bayesianas para Classificação de Padrões

- **Rede única para todas as classes:**
  - » **Conjunto de variáveis que descrevem os padrões, *V***
  - » **variável que representa a classe *j***
  - » **outras variáveis que representam efeitos físicos importantes**
  - » **O método de inferência calcula *P(j|V)***
  - » ***Vantagem:* o modelo faz uso eficiente das semelhanças e diferenças entre todas as classes**

- **Abordagem múltiplo modelo:**
  - » **definir um modelo separado para cada classe *j***
  - » **cada modelo é rotulado de acordo com o número da classe: a rede *j* representa a distribuição**

$$P(V, h_j \mid j)$$

$$P(V \mid j) = \sum_{h_j} P(V, h_j \mid j)$$

$$h_j \equiv \text{atributos "escondidos" para a classe } j$$

  - » **Classificação usando a regra de Bayes:**

$$\hat{P}(j \mid V) = \frac{P(V \mid j)P(j)}{\sum_i P(V \mid i)P(i)}$$

$$j = \arg\max \hat{P}(j \mid V)$$

Aprendizagem de Redes Bayesianas

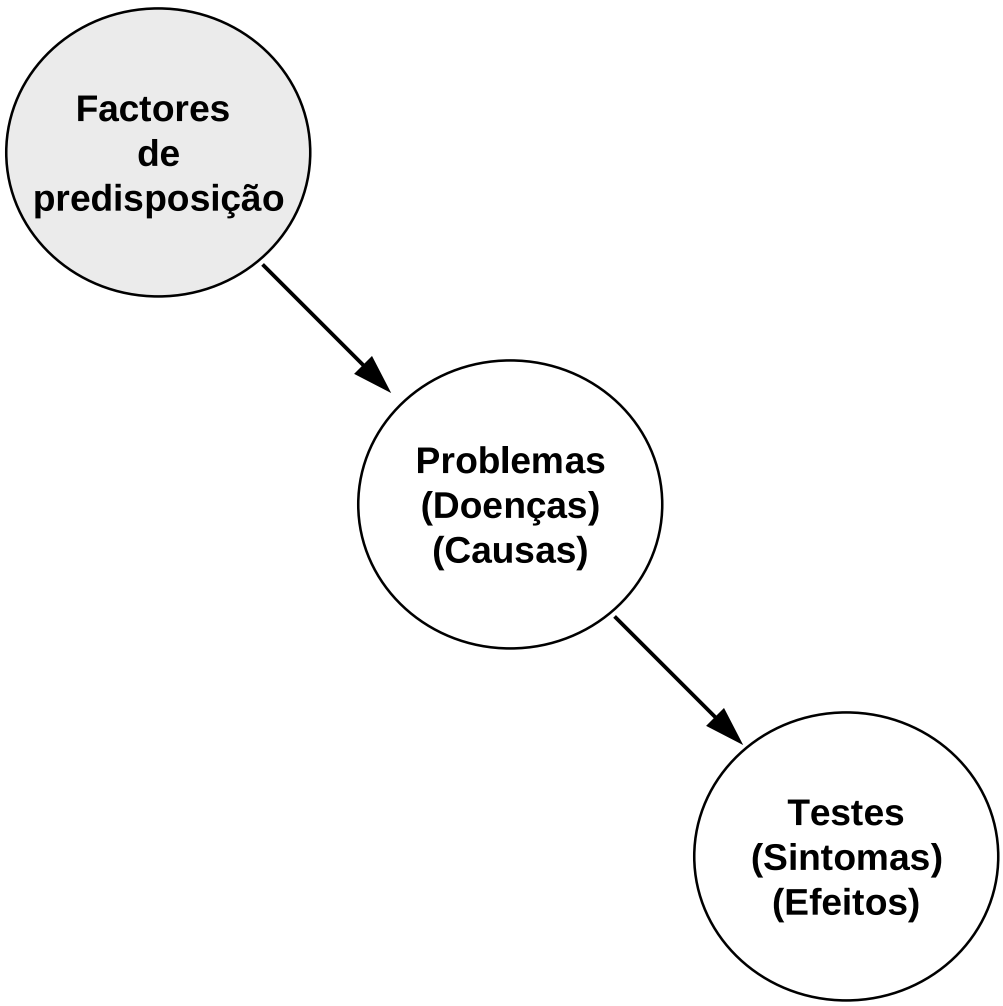
- **Distribuições estatísticas**
- **Topologia + estatísticas**
  - » **Algoritmos genéticos**
  - » **algoritmos estocásticos**

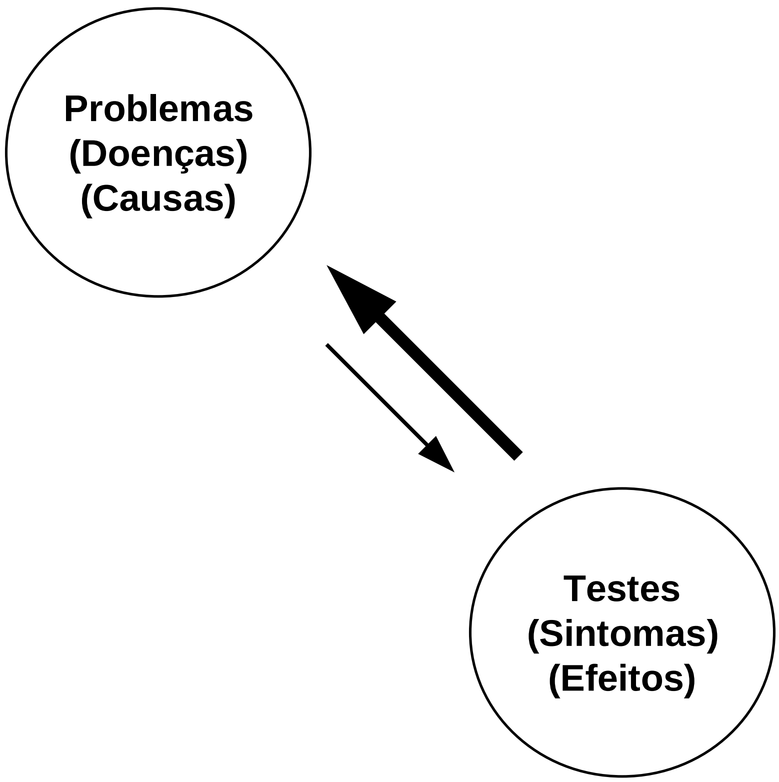
Software de Redes Bayesianas

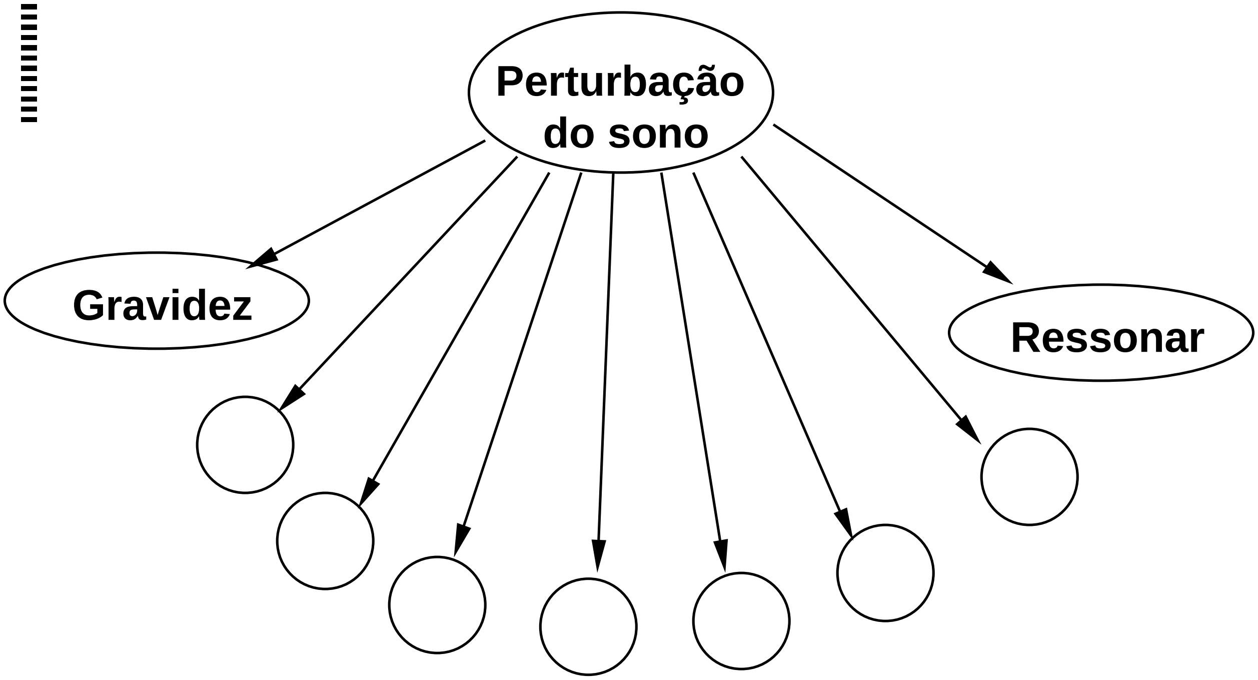
- **HUGIN (Andersen et al. 1995)**
  - » **www.hugin.dk/lat-bn.html**

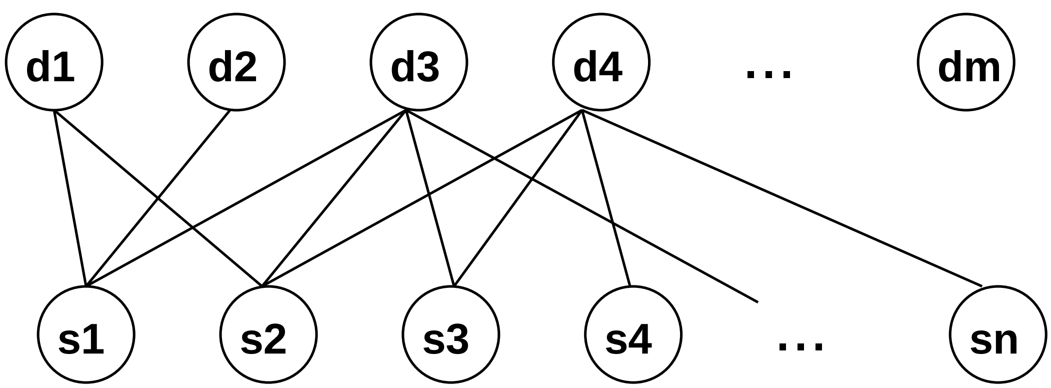
- **IDEAL (srinvas & Breese, 1990)**
- **PRESS (Gammerman et al, 1995)**

Uma rede Bayesiana para diagnóstico




Problemas
(Doenças)
(Causas)
Testes
(Sintomas)
(Efeitos)


Perturbação do sono
Gravidez
Ressonar


d1
d2
d3
d4
...
dm
s1
s2
s3
s4
...
sn