Inférence dans les réseaux bayésiens

Christophe Gonzales

LIP6 – Université Paris 6, France

Plan du cours 2

1. L'algorithme de Shafer-Shenoy
2. D'autres algorithmes d'inférence
3. Construction d'un arbre de jonction

Une présentation unifiée des algorithmes d'inférence

Définition

Inférence : calculer des probabilités (marginales, a priori, a posteriori, etc.).

Les algorithmes d'inférence :

- *Lazy propagation*
- *Shafer−Shenoy*
- *Jensen (HUGIN)*

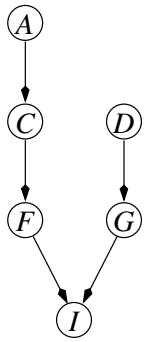
} méthodes non orientées

- *Pearl*

} méthode orientée

Osons le calcul des probabilités a priori

$$P(A, C, D, F, G, I) = P(A)P(C|A)P(F|C)P(D)P(G|D)P(I|F, G)$$



Calcul de $P(I)$?

Shafer-Shenoy brut de fonderie

$$P(A,C,D,F,G,I) = P(A)P(C|A)P(F|C)P(D)P(G|D)P(I|F,G)$$

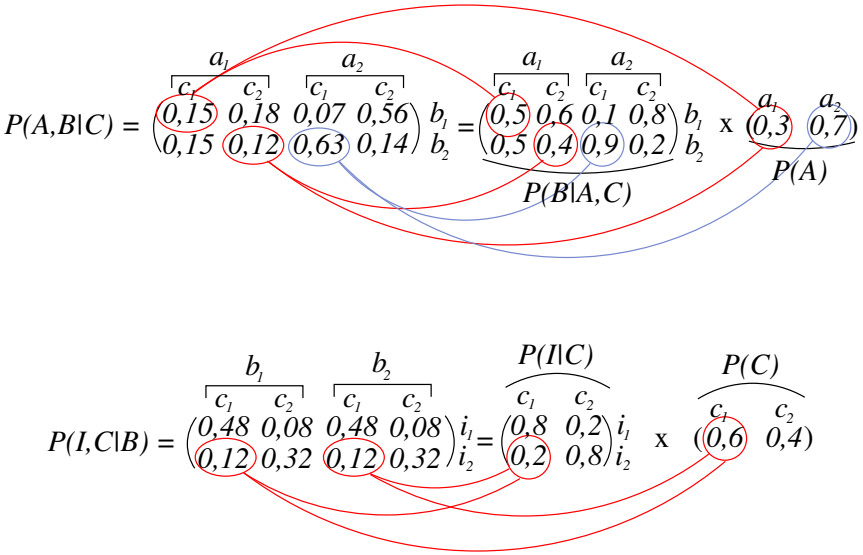
$$P(I) = \sum_G \left(\sum_F \left(\sum_D \left(\sum_C \left(\sum_A P(A,C,D,F,G,I) \right) \right) \right) \right)$$

$$\sum_A P(A,C,D,F,G,I) = \underbrace{\left(\sum_A P(A)P(C|A) \right)}_{P(C)} P(F|C)P(D)P(G|D)P(I|F,G)$$

$$\sum_A P(A,C,D,F,G,I) = P(C)P(F|C)P(D)P(G|D)P(I|F,G)$$

$$\sum_C \sum_A P(A,C,D,F,G,I) = \underbrace{\left(\sum_C P(C)P(F|C) \right)}_{P(F)} P(D)P(G|D)P(I|F,G)$$

Dissection du produit de deux probabilités



Shafer-Shenoy graphique (1/6)

Séquence d'élimination

A C D F G

$$P(A,C,D,F,G,I) = P(A)P(C|A)\,P(F|C)P(D)P(G|D)P(I|F,G)$$

A	AC	FC	D	GD	IFG
$P(A)$	$P(C\|A)$	$P(F\|C)$	$P(D)$	$P(G\|D)$	$P(I\|F,G)$

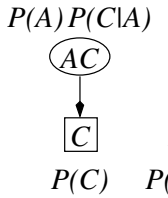
somme sur $A \Longrightarrow P(C) = \sum_A P(A)P(C|A)$

Shafer-Shenoy graphique (2/6)

Séquence d'élimination

A C D F G

$$P(C,D,F,G,I) = P(C)P(F|C)\,P(D)P(G|D)P(I|F,G)$$



FC	D	GD	IFG
$P(F\|C)$	$P(D)$	$P(G\|D)$	$P(I\|F,G)$

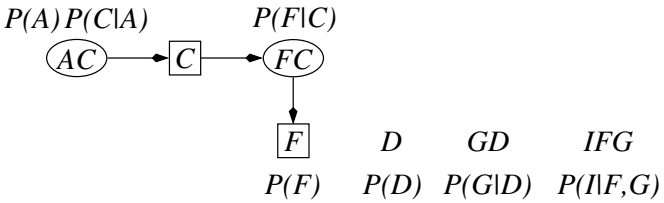
somme sur $C \Longrightarrow P(F) = \sum_C P(C)P(F|C)$

Shafer-Shenoy graphique (3/6)

Séquence d'élimination

$A\ C\ \boxed{D}\ F\ G$

$P(D,F,G,I) = P(F)\ \boxed{P(D)P(G|D)}\ P(I|F,G)$



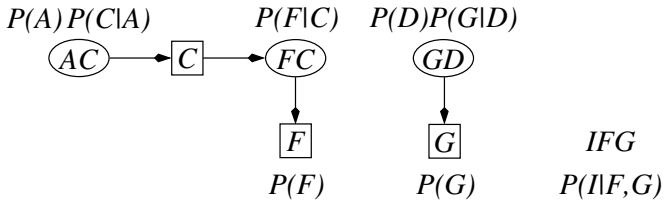
somme sur $D \Longrightarrow P(G) = \sum_D P(D)P(G|D)$

Shafer-Shenoy graphique (4/6)

Séquence d'élimination

$A\ C\ D\ \boxed{F}\ G$

$P(D,F,G,I) = \boxed{P(F)P(I|F,G)}\ P(G)$



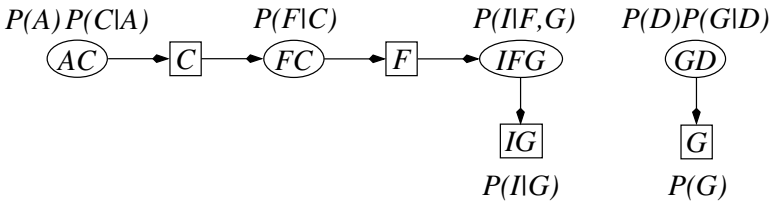
somme sur $F \Longrightarrow P(I|G) = \sum_F P(F)P(I|F,G)$

Shafer-Shenoy graphique (5/6)

Séquence d'élimination

$A\ C\ D\ F\ \boxed{G}$

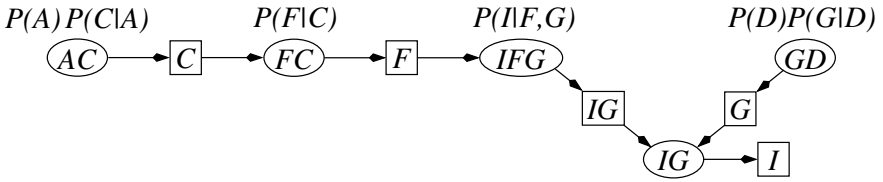
$P(D,F,G,I) = \boxed{P(I|G)P(G)}$



somme sur $G \Longrightarrow P(I) = \sum_G P(G)P(I|G)$

Shafer-Shenoy graphique (6/6)

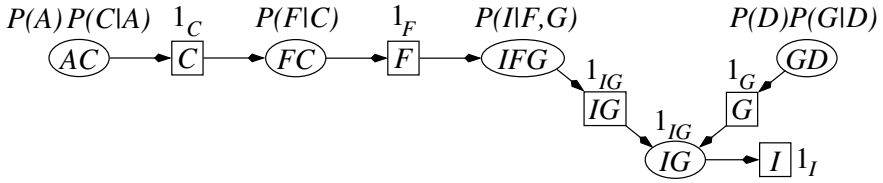
Le graphe final obtenu par Shafer-Shenoy



Algorithme de Shafer-Shenoy

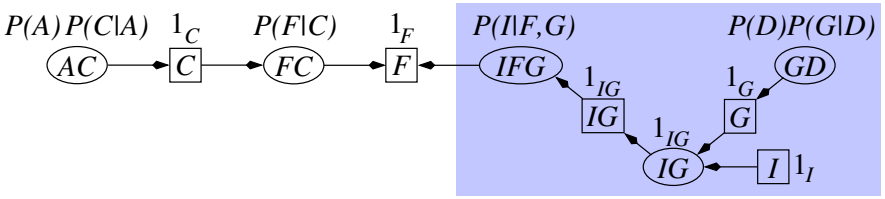
1. Se donner une séquence d'élimination des nœuds $\Longrightarrow$ *join tree*,
2. propager les impacts dans le sens des flèches :
   - dans les ellipses (cliques), on effectue des multiplications,
   - dans les rectangles (séparateurs), on effectue des additions (projections).

Quelques remarques sur Shafer-Shenoy (1/3)

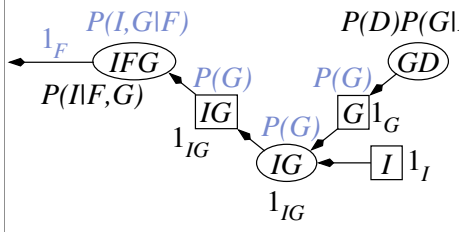


1. Loi jointe = produit des fonctions des cliques et des séparateurs : $P(A, C, D, F, G, I) = P(A)P(C|A)P(F|C)P(D)P(G|D)P(I|F, G)$.
2. Élimination récursive des cliques et séparateurs « externes » $\Longrightarrow$ le produit des fonctions des cliques et des séparateurs restants = loi jointe des variables restantes.

Quelques remarques sur Shafer-Shenoy (2/3)



$P(A) \times P(C|A) \times P(F|C) \times \mathbb{1}_C \times \mathbb{1}_F = P(A, C, F)$.



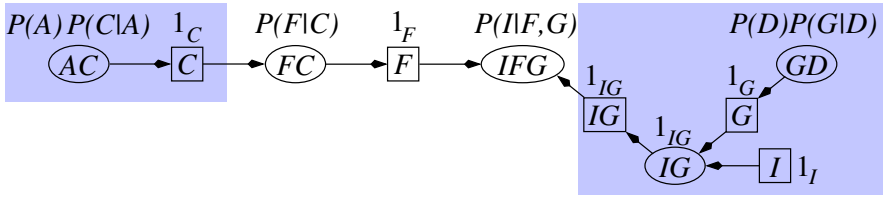
séparateur G : $\sum_D P(D)P(G|D)$

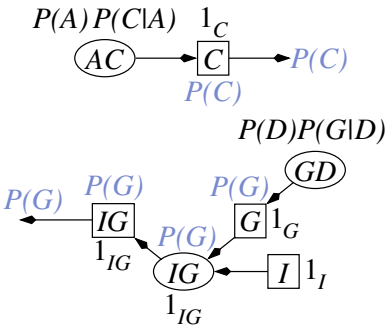
clique IG : $P(G) \times \mathbb{1}_{IG} \times \mathbb{1}_G \times \mathbb{1}_I$

clique IFG : $\sum_{IG} P(I|F, G) \times P(G)$

$\Longrightarrow$ en sortie de IFG : $\mathbb{1}_F$

Quelques remarques sur Shafer-Shenoy (3/3)





Produit des cliques et séparateurs restants :

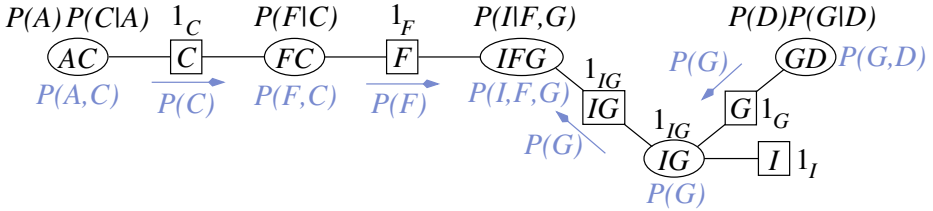
$P(C)\ P(F|C)\mathbb{1}_F P(I|F, G)\ P(G)$

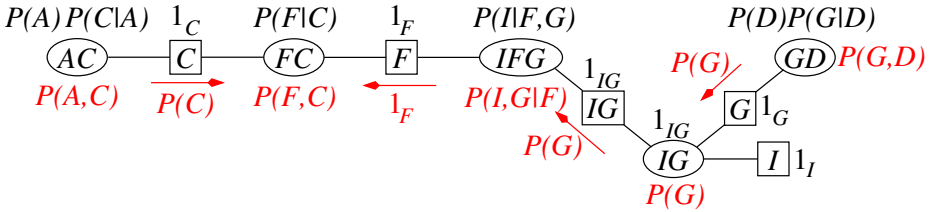
$= P(I, C, F, G)$

Conclusion : on peut utiliser le même graphe pour calculer toutes les probabilités marginales

C'est pas le deux en un, mais le tout en deux (1/3)

En bleu : les calculs de $P(I, F, G)$, en rouge, ceux de $P(F, C)$





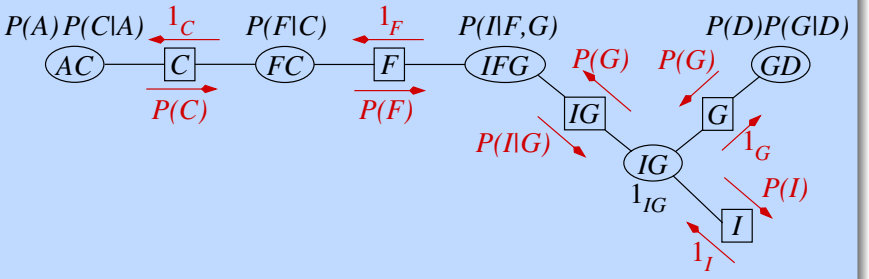
C'est pas le deux en un, mais le tout en deux (2/3)

Algorithme de Shafer-Shenoy

1. Chaque séparateur contient deux messages initialisés à $\mathbb{1}$, un en direction de chaque clique voisine.
2. Chaque nœud du join tree envoie des messages vers ses voisins en respectant les deux règles suivantes :
   1. avant d'envoyer un message vers son voisin X, le nœud Y attend que tous ses autres voisins lui aient envoyé leur message.
   2. le message d'un nœud Y vers son voisin X est le produit de tous les messages reçus par Y, à l'exception de celui envoyé par X, et de la table stockée par Y, le tout marginalisé sur X (c'est-à-dire sommé sur les variables de $Y \setminus X$).

C'est pas le deux en un, mais le tout en deux (3/3)

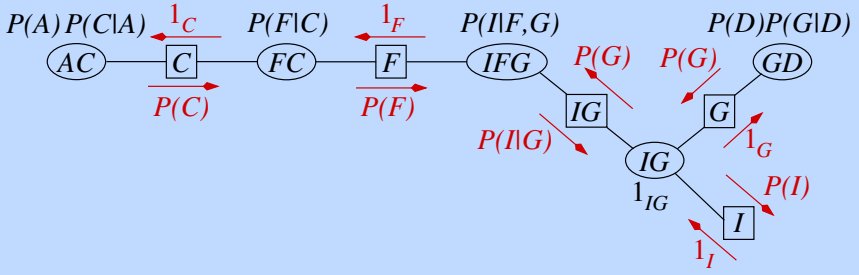
Algorithme de Shafer-Shenoy



À la fin de l'algorithme, pour tout nœud X, le produit de la table stockée en X par l'ensemble des messages envoyés à X est la probabilité jointe des variables de X.

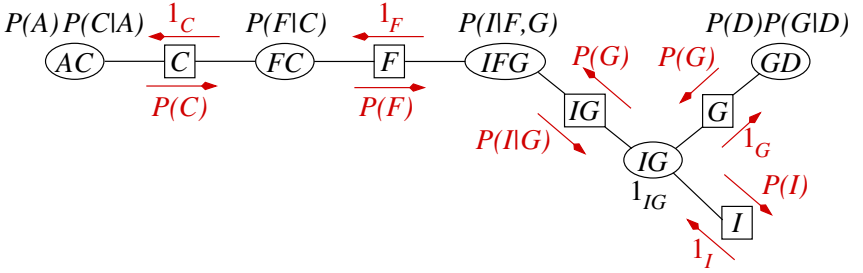
Shafer-Shenoy : ce qu'il faut retenir

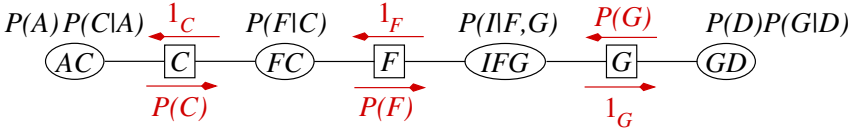
Algorithme de Shafer-Shenoy



- cliques = ellipses, séparateurs = rectangles
- algorithme par envoi de messages
- opérateurs = $+$ sur les séparateurs, $\times$ sur les cliques

Les arbres de jonction





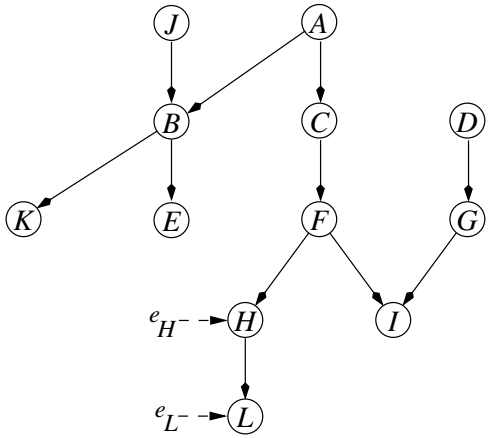
arbre de jonction : suppression des cliques incluses dans d'autres cliques

Soyons observateur : les probabilités a posteriori

Nous observons des informations e_H sur H et e_L sur L.

On veut connaître $P(I|e_H, e_L)$

$\Longrightarrow$ on va calculer $P(I, e_H, e_L)$ car c'est plus simple



Quelles sont ces informations ?

Teneur des informations

informations = observations :

e_L = « L ne peut plus prendre les valeurs l_1 et l_4 »

Entrée de ces observations : $P(e_L|L)$

Vous avez dit calcul de $P(e_L|L)$? (1/2)

observation : e_L = « Le capteur m'indique que L ne peut plus prendre les valeurs l_1 et l_4 »

$$P(e_L|L) = \begin{array}{|c|l} \hline 0 & l_1 \\ \hline 1 & l_2 \\ \hline 1 & l_3 \\ \hline 0 & l_4 \\ \hline \end{array}$$

⚠ dimension de $P(e_L|L) = |L|$

Vous avez dit calcul de $P(e_L|L)$? (2/2)

observation : e_L = « Le capteur m'indique que L ne peut plus prendre les valeurs l_1 et l_4. Mais je n'ai pas totalement confiance en ce capteur. Je pense qu'il y a 10% de chances pour que le capteur se trompe au sujet de l_1 et 20% au sujet au sujet de l_3. »

$$P(e_L|L) = \begin{array}{|c|l} \hline 0.1 & l_1 \\ \hline 1 & l_2 \\ \hline 0.8 & l_3 \\ \hline 0 & l_4 \\ \hline \end{array}$$

Hypothèses et conséquences (1/4)

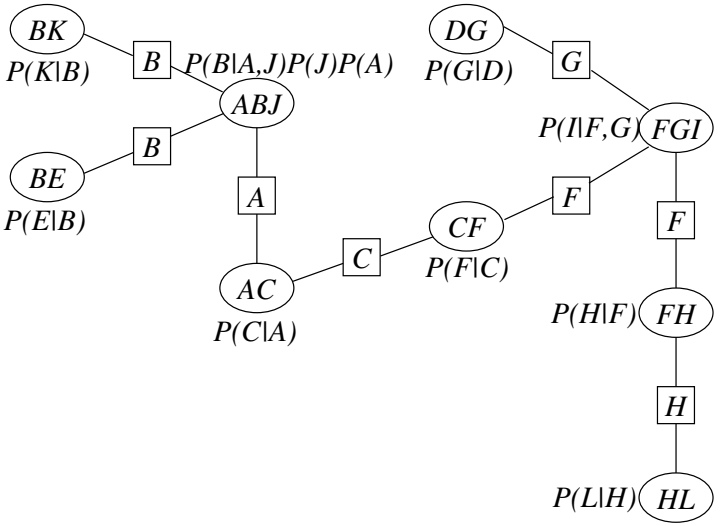
Hypothèse

Toute information e_X sur un nœud X est indépendante du reste du réseau bayésien conditionnellement à X.

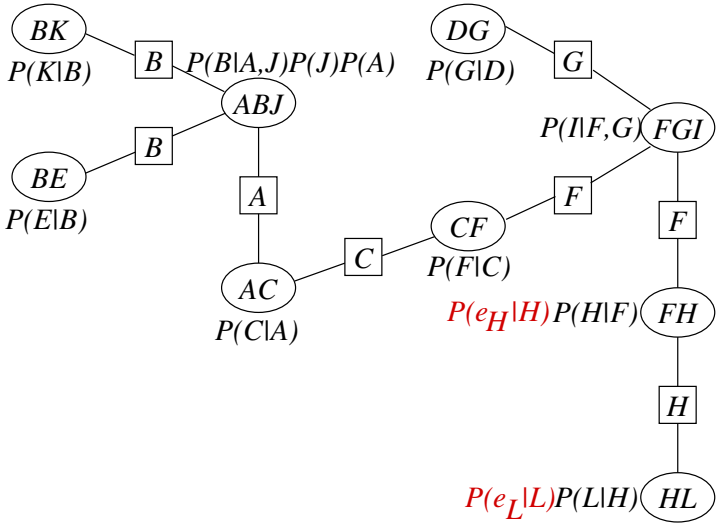
Conséquence

$$P(A,B,C,D,E,F,G,H,I,J,K,L,e_H,e_L) = P(e_H|A,B,C,D,E,F,G,H,I,J,K,L,e_L)\times P(e_L|A,B,C,D,E,F,G,H,I,J,K,L)\times P(A,B,C,D,E,F,G,H,I,J,K,L)$$
$$= P(e_H|H)P(e_L|L)P(A,B,C,D,E,F,G,H,I,J,K,L)$$

Hypothèses et conséquences (2/4)



Hypothèses et conséquences (3/4)



Hypothèses et conséquences (4/4)

L'algorithme de Shafer-Shenoy permet donc de calculer dans la clique FGI : $P(F,G,I,e_H,e_L)$

$$\Longrightarrow P(I,e_H,e_L) = \sum_{F,G} P(F,G,I,e_H,e_L)$$

$$\Longrightarrow P(I|e_H,e_L) = \frac{P(I,e_H,e_L)}{P(e_H,e_L)}$$

Or $P(e_H,e_L) = \sum_I P(I,e_H,e_L)$

Donc $P(I|e_H,e_L) = \dfrac{P(I,e_H,e_L)}{\sum_I P(I,e_H,e_L)}$.

Résumé sur l'algorithme de Shafer-Shenoy

Algorithme de Shafer-Shenoy

Pour calculer une proba *a posteriori* $P(Y|e)$:

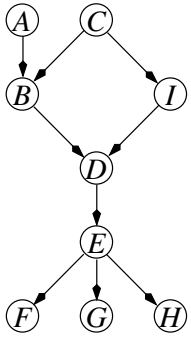
1. construire l'arbre de jonction
2. insérer les probas conditionnelles du réseau bayésien dans les cliques
3. insérer des tables contenant uniquement des 1 dans les séparateurs
4. pour chaque information e_X, calculer $P(e_X|X)$ et insérer cette proba dans une clique contenant X
5. envoyer les messages dans l'arbre de jonction
6. choisir une clique contenant Y, faire le produit de sa table de proba par tous les messages qui lui ont été envoyés, puis sommer sur toutes les variables $\neq Y \Longrightarrow P(Y, e)$
7. normaliser $P(Y, e) \Longrightarrow P(Y|e)$

Petit guide pratique sur Shafer-Shenoy

Quelques références

1. G. Shafer (1996) *Probabilistic expert systems*, Society for Industrial and Applied Mathematics.
2. P.P. Shenoy, G. Shafer (1990) *Axioms for probability and belief-function propagation*, Uncertainty in Artificial Intelligence 4, pp.169–198.
3. P.P. Shenoy (1997) *Binary join trees for computing marginals in the Shenoy-Shafer architecture*, International Journal of Approximate Reasoning 17, pp.1–25.
4. V. Lepar, P.P. Shenoy (1998) *A Comparison of Lauritzen-Spiegelhalter, Hugin and Shenoy-Shafer Architectures for Computing Marginals of Probability Distributions*, Proceedings of UAI-98, pp.328–337.

Exemple à l'usage des étudiants studieux (1/9)



$$P(A) = \begin{matrix} a_1 & a_2 \end{matrix} \begin{pmatrix} 0{,}3 & 0{,}7 \end{pmatrix}$$

$$P(B|A,C) = \begin{matrix} & \overbrace{}^{a_1} & & \overbrace{}^{a_2} \\ & c_1 & c_2 & c_1 & c_2 \end{matrix} \begin{matrix} \\ b_1 \\ b_2 \end{matrix} \begin{pmatrix} 0{,}5 & 0{,}6 & 0{,}1 & 0{,}8 \\ 0{,}5 & 0{,}4 & 0{,}9 & 0{,}2 \end{pmatrix}$$

$$P(D|B,I) = \begin{matrix} & \overbrace{}^{b_1} & & \overbrace{}^{b_2} \\ & i_1 & i_2 & i_1 & i_2 \end{matrix} \begin{matrix} \\ d_1 \\ d_2 \end{matrix} \begin{pmatrix} 0{,}1 & 0{,}4 & 0{,}7 & 0{,}8 \\ 0{,}9 & 0{,}6 & 0{,}3 & 0{,}2 \end{pmatrix}$$

$$P(F|E) = \begin{matrix} e_1 & e_2 & e_3 \end{matrix} \begin{pmatrix} 0{,}2 & 0{,}7 & 0{,}4 \\ 0{,}8 & 0{,}3 & 0{,}6 \end{pmatrix} \begin{matrix} f_1 \\ f_2 \end{matrix}$$

$$P(H|E) = \begin{matrix} e_1 & e_2 & e_3 \end{matrix} \begin{pmatrix} 0{,}5 & 0{,}6 & 0{,}2 \\ 0{,}5 & 0{,}4 & 0{,}8 \end{pmatrix} \begin{matrix} h_1 \\ h_2 \end{matrix}$$

$$P(C) = \begin{matrix} c_1 & c_2 \end{matrix} \begin{pmatrix} 0{,}6 & 0{,}4 \end{pmatrix}$$

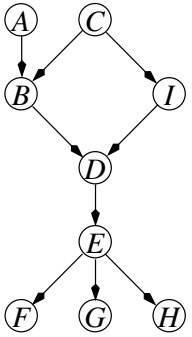
$$P(I|C) = \begin{matrix} c_1 & c_2 \end{matrix} \begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}2 & 0{,}8 \end{pmatrix} \begin{matrix} i_1 \\ i_2 \end{matrix}$$

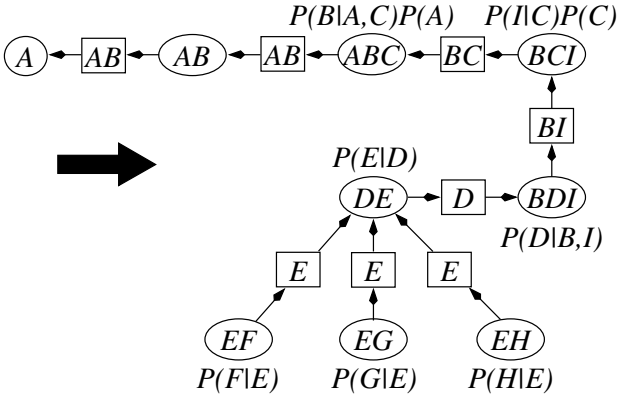
$$P(E|D) = \begin{matrix} d_1 & d_2 \end{matrix} \begin{pmatrix} 0{,}4 & 0{,}3 \\ 0{,}5 & 0{,}4 \\ 0{,}1 & 0{,}3 \end{pmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \end{matrix}$$

$$P(G|E) = \begin{matrix} e_1 & e_2 & e_3 \end{matrix} \begin{pmatrix} 0{,}6 & 0{,}1 & 0{,}3 \\ 0{,}4 & 0{,}9 & 0{,}7 \end{pmatrix} \begin{matrix} g_1 \\ g_2 \end{matrix}$$

Exemple à l'usage des étudiants studieux (2/9)

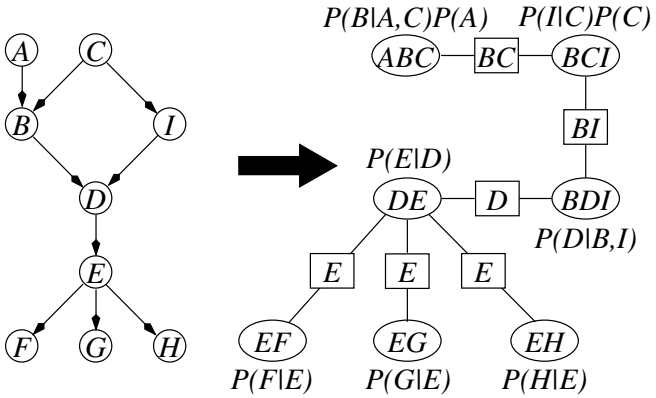
Séquence d'élimination : $F, H, G, E, D, I, C, B, A$



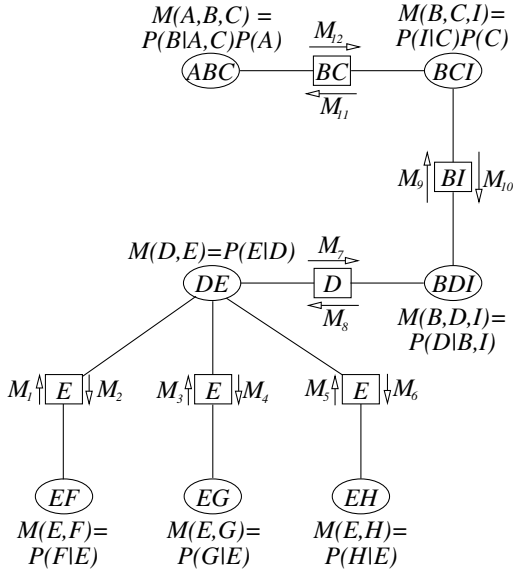


Exemple à l'usage des étudiants studieux (3/9)

Séquence d'élimination : $F, H, G, E, D, I, C, B, A$



Exemple à l'usage des étudiants studieux (4/9)



initialisation

$$M(A,B,C) = \begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} a_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}15 & 0{,}18 \\ 0{,}15 & 0{,}12 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} a_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}07 & 0{,}56 \\ 0{,}63 & 0{,}14 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}$$

$$M(B,C,I) = \begin{array}{c} \\ i_1 \\ i_2 \end{array}\begin{array}{c} \begin{array}{cc} b_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} b_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 \end{pmatrix} \end{array}\begin{array}{c} \\ i_1 \\ i_2 \end{array}$$

$$M_1 = M_2 = M_3 = M_4 = M_5 = M_6 = \begin{array}{cccc} e_1 & e_2 & e_3 & \\ (1 & 1 & 1) & \end{array}$$

$$M_7 = M_8 = \begin{array}{cc} d_1 & d_2 \\ (1 & 1) \end{array}$$

$$M_9 = M_{10} = \begin{array}{c} \\ i_1 \\ i_2 \end{array}\begin{array}{c} \begin{array}{cc} i_1 & i_2 \end{array} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}$$

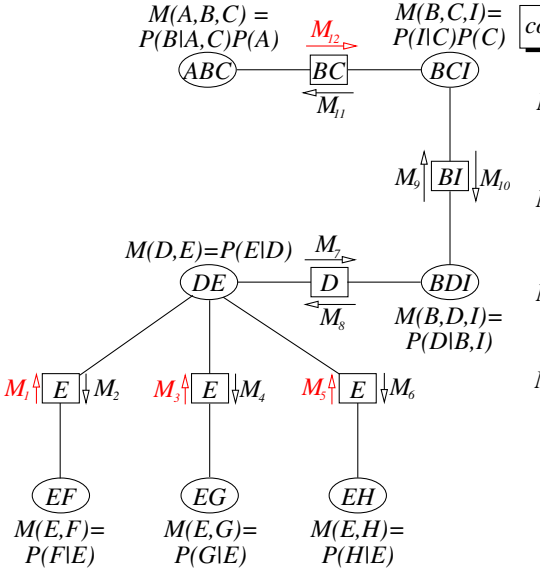
$$M_{11} = M_{12} = \begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}$$

Rappel : dissection du produit de deux probabilités

$$M(A,B,C) = \begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} a_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}15 & 0{,}18 \\ 0{,}15 & 0{,}12 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} a_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}07 & 0{,}56 \\ 0{,}63 & 0{,}14 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array} = \underbrace{\begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} a_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}5 & 0{,}6 \\ 0{,}5 & 0{,}4 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} a_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}1 & 0{,}8 \\ 0{,}9 & 0{,}2 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}}_{P(B|A,C)} \times \underbrace{\begin{array}{cc} a_1 & a_2 \\ (0{,}3 & 0{,}7) \end{array}}_{P(A)}$$

$$M(B,C,I) = \begin{array}{c} \\ i_1 \\ i_2 \end{array}\begin{array}{c} \begin{array}{cc} b_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} b_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 \end{pmatrix} \end{array}\begin{array}{c} \\ i_1 \\ i_2 \end{array} = \underbrace{\begin{array}{c} \\ i_1 \\ i_2 \end{array}\begin{array}{c} \begin{array}{cc} i_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}2 & 0{,}8 \end{pmatrix} \end{array}\begin{array}{c} \\ i_1 \\ i_2 \end{array}}_{P(I|C)} \times \underbrace{\begin{array}{cc} c_1 & c_2 \\ (0{,}6 & 0{,}4) \end{array}}_{P(C)}$$

Exemple à l'usage des étudiants studieux (5/9)



collecte: les feuilles envoient des messages

$$M_1 := \mathrm{Proj}_E M(E,F) = \mathrm{Proj}_E \begin{array}{cccc} & e_1 & e_2 & e_3 \\ \begin{matrix} f_1 \\ f_2 \end{matrix} & \begin{pmatrix} 0{,}2 & 0{,}7 & 0{,}4 \\ 0{,}8 & 0{,}3 & 0{,}6 \end{pmatrix} \end{array}$$

$$:= \begin{array}{cccc} & e_1 & e_2 & e_3 \\ & (1 & 1 & 1) \end{array}$$

$$M_3 := \mathrm{Proj}_E M(E,G) = \mathrm{Proj}_E \begin{array}{cccc} & e_1 & e_2 & e_3 \\ \begin{matrix} g_1 \\ g_2 \end{matrix} & \begin{pmatrix} 0{,}6 & 0{,}1 & 0{,}3 \\ 0{,}4 & 0{,}9 & 0{,}7 \end{pmatrix} \end{array}$$

$$:= \begin{array}{cccc} & e_1 & e_2 & e_3 \\ & (1 & 1 & 1) \end{array}$$

$$M_5 := \mathrm{Proj}_E M(E,H) = \mathrm{Proj}_E \begin{array}{cccc} & e_1 & e_2 & e_3 \\ \begin{matrix} h_1 \\ h_2 \end{matrix} & \begin{pmatrix} 0{,}5 & 0{,}6 & 0{,}2 \\ 0{,}5 & 0{,}4 & 0{,}8 \end{pmatrix} \end{array}$$

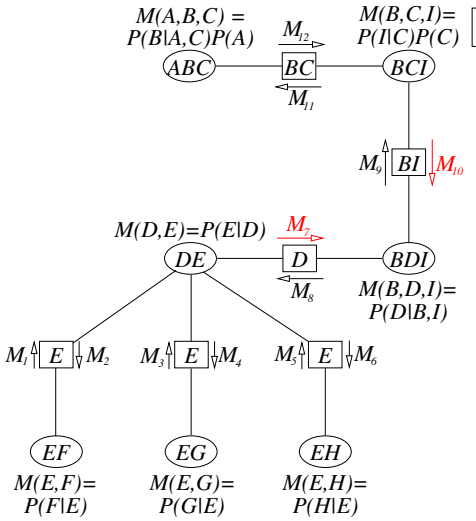
$$:= \begin{array}{cccc} & e_1 & e_2 & e_3 \\ & (1 & 1 & 1) \end{array}$$

$$M_{12} := \mathrm{Proj}_{BC} M(A,B,C)$$

$$:= \mathrm{Proj}_{BC} \begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} a_1 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}15 & 0{,}18 \\ 0{,}15 & 0{,}12 \end{pmatrix} \end{array}\begin{array}{c} \begin{array}{cc} a_2 & \\ c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}07 & 0{,}56 \\ 0{,}63 & 0{,}14 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}$$

$$:= \begin{array}{c} \\ b_1 \\ b_2 \end{array}\begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \begin{pmatrix} 0{,}22 & 0{,}74 \\ 0{,}78 & 0{,}26 \end{pmatrix} \end{array}\begin{array}{c} \\ b_1 \\ b_2 \end{array}$$

Exemple à l'usage des étudiants studieux (6/9)



collecte (suite)

$$M_7 := \mathrm{Proj}_D(M(D,E) \times M_1 \times M_3 \times M_5)$$

$$:= \mathrm{Proj}_D \begin{pmatrix} d_1 & d_2 \\ 0{,}4 & 0{,}3 \\ 0{,}5 & 0{,}4 \\ 0{,}1 & 0{,}3 \end{pmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \end{matrix} = \begin{matrix} d_1 & d_2 \\ (1 & 1) \end{matrix}$$

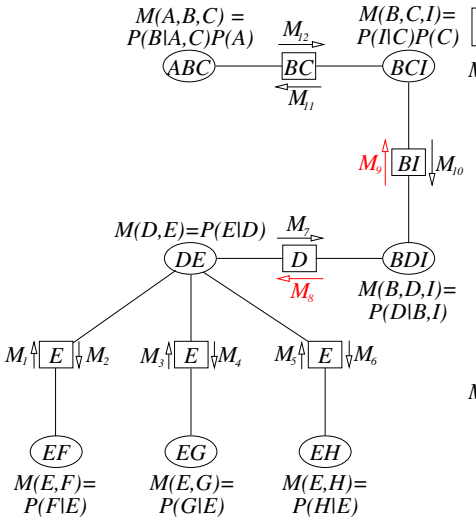
$$M_{10} := \mathrm{Proj}_{BI}(M(B,C,I) \times M_{12})$$

$$:= \mathrm{Proj}_{BI} \left[\begin{matrix} \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}48 & 0{,}08 & 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 & 0{,}12 & 0{,}32 \end{pmatrix} \begin{matrix} i_1 \\ i_2 \end{matrix} \end{matrix} \times \begin{matrix} c_1 & c_2 \\ \begin{pmatrix} 0{,}22 & 0{,}74 \\ 0{,}78 & 0{,}26 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix} \right]$$

$$:= \mathrm{Proj}_{BI} \begin{matrix} \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}1056 & 0{,}0592 & 0{,}3744 & 0{,}0208 \\ 0{,}0264 & 0{,}2368 & 0{,}0936 & 0{,}0832 \end{pmatrix} \begin{matrix} i_1 \\ i_2 \end{matrix} \end{matrix}$$

$$:= \begin{matrix} b_1 & b_2 \\ \begin{pmatrix} 0{,}1648 & 0{,}3952 \\ 0{,}2632 & 0{,}1746 \end{pmatrix} \end{matrix} \begin{matrix} i_1 \\ i_2 \end{matrix}$$

Exemple à l'usage des étudiants studieux (7/9)



début de la diffusion

$$M_8 := \mathrm{Proj}_D(M(B,D,I) \times M_{10})$$

$$:= \mathrm{Proj}_D \left[\begin{matrix} \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}1 & 0{,}4 & 0{,}7 & 0{,}8 \\ 0{,}9 & 0{,}6 & 0{,}3 & 0{,}2 \end{pmatrix} \begin{matrix} d_1 \\ d_2 \end{matrix} \end{matrix} \times \begin{matrix} b_1 & b_2 \\ \begin{pmatrix} 0{,}1648 & 0{,}3952 \\ 0{,}2632 & 0{,}1746 \end{pmatrix} \end{matrix} \begin{matrix} i_1 \\ i_2 \end{matrix} \right]$$

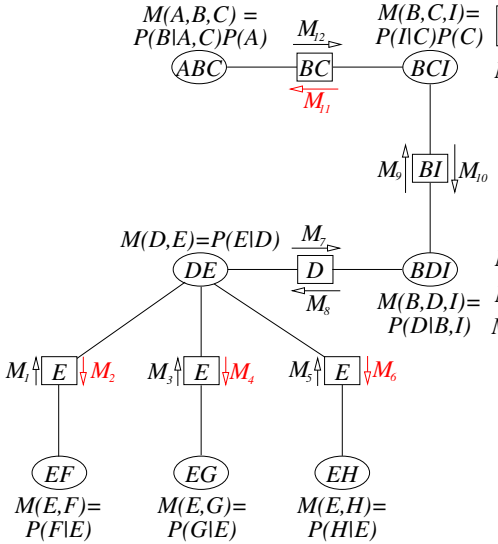
$$:= \mathrm{Proj}_D \begin{matrix} \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}01648 & 0{,}10528 & 0{,}27664 & 0{,}13968 \\ 0{,}14832 & 0{,}15792 & 0{,}11856 & 0{,}03492 \end{pmatrix} \begin{matrix} d_1 \\ d_2 \end{matrix} \end{matrix}$$

$$:= \begin{pmatrix} 0{,}53808 \\ 0{,}45972 \end{pmatrix} \begin{matrix} d_1 \\ d_2 \end{matrix}$$

$$M_9 := \mathrm{Proj}_{BI}(M(B,D,I) \times M_7)$$

$$:= \mathrm{Proj}_{BI} \left[\begin{matrix} \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} i_1 & i_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}1 & 0{,}4 & 0{,}7 & 0{,}8 \\ 0{,}9 & 0{,}6 & 0{,}3 & 0{,}2 \end{pmatrix} \begin{matrix} d_1 \\ d_2 \end{matrix} \end{matrix} \times \begin{matrix} d_1 & d_2 \\ (1 & 1) \end{matrix} \right] = \begin{matrix} i_1 & i_2 \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix}$$

Exemple à l'usage des étudiants studieux (8/9)



fin de la diffusion

$$M_2 := \mathrm{Proj}_E(M(D,E) \times M_3 \times M_5 \times M_8)$$

$$:= \mathrm{Proj}_E \left[\begin{matrix} d_1 & d_2 \\ \begin{pmatrix} 0{,}4 & 0{,}3 \\ 0{,}5 & 0{,}4 \\ 0{,}1 & 0{,}3 \end{pmatrix} \begin{matrix} e_1 \\ e_2 \\ e_3 \end{matrix} \end{matrix} \times \begin{pmatrix} 0{,}53808 \\ 0{,}45972 \end{pmatrix} \begin{matrix} d_1 \\ d_2 \end{matrix} \right]$$

$$:= \begin{matrix} e_1 & e_2 & e_3 \\ (0{,}35314 & 0{,}45292 & 0{,}19171) \end{matrix}$$

$$M_4 := \mathrm{Proj}_E(M(D,E) \times M_1 \times M_5 \times M_8) := M_2$$

$$M_6 := \mathrm{Proj}_E(M(D,E) \times M_1 \times M_3 \times M_8) := M_2$$

$$M_{11} := \mathrm{Proj}_{BC}(M(B,C,I) \times M_9)$$

$$:= \mathrm{Proj}_{BC} \left[\begin{matrix} \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_1} & \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{b_2} \\ \begin{pmatrix} 0{,}48 & 0{,}08 & 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 & 0{,}12 & 0{,}32 \end{pmatrix} \begin{matrix} i_1 \\ i_2 \end{matrix} \end{matrix} \times \begin{matrix} i_1 & i_2 \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix} \right]$$

$$:= \begin{matrix} c_1 & c_2 \\ \begin{pmatrix} 0{,}60 & 0{,}40 \\ 0{,}60 & 0{,}40 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix}$$

Exemple à l'usage des étudiants studieux (9/9)

$$M_{11} \times M(A,B,C) = \begin{matrix} c_1 & c_2 \\ \begin{pmatrix} 0{,}60 & 0{,}40 \\ 0{,}60 & 0{,}40 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix} \times \begin{matrix} \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{a_1} & \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{a_2} \\ \begin{pmatrix} 0{,}15 & 0{,}18 & 0{,}07 & 0{,}56 \\ 0{,}15 & 0{,}12 & 0{,}63 & 0{,}14 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix}$$

$$= \begin{matrix} \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{a_1} & \overbrace{\begin{matrix} c_1 & c_2 \end{matrix}}^{a_2} \\ \begin{pmatrix} 0{,}090 & 0{,}072 & 0{,}042 & 0{,}224 \\ 0{,}090 & 0{,}048 & 0{,}378 & 0{,}056 \end{pmatrix} \end{matrix} \begin{matrix} b_1 \\ b_2 \end{matrix}$$

1. Petit exercice de calcul mental : calculez la projection de la matrice ci-dessus sur *BC*.
2. Petit exercice de déduction : à quoi correspond cette matrice ?
3. Ultime exercice à l'usage de l'élite (et des biens nantis) : calculez le produit $M_{11} \times M_{12}$.
4. Un dernier pour la route : dites « bizarre, bizarre, comme c'est bizarre », et expliquez pourquoi c'est bizarre.

❷ Existe-t-il d'autres

algorithmes d'inférence ?

De Shafer-Shenoy à Lazy propagation

si $P(A,B,C,D,E) = P(A)P(B)P(C|A,B)P(D)P(E|C,D)$

alors $P(B,C,D,E) = \sum_A P(A)P(B)P(C|A,B)P(D)P(E|C,D)$

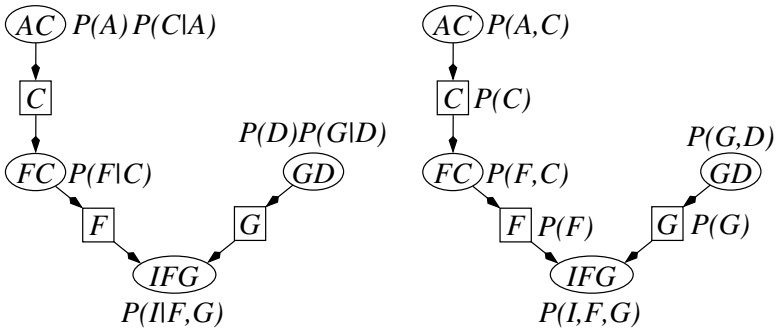
$$= \left(\sum_A P(A)P(C|A,B)\right) P(B)P(D)P(E|C,D)$$

$\Longrightarrow$ on a intérêt à ne pas effectuer les produits avant les sommes

Lazy propagation

Principe : garder les produits sous forme de listes et n'effectuer les multiplications que lorsque c'est nécessaire.

De Shafer-Shenoy à Jensen



Shafer-Shenoy : élimination de $A \Longrightarrow P(C, e_A) = \sum_A P(A, e_A)P(C|A)$

élimination de $F \Longrightarrow P(F, C, e_A) = P(F|C)P(C, e_A)$

Jensen : élimination de $F \Longrightarrow P(F, C, e_A) = \dfrac{P(F,C)}{P(C)} P(C, e_A)$

Petit guide sur Jensen et Lazy Propagation

Quelques références

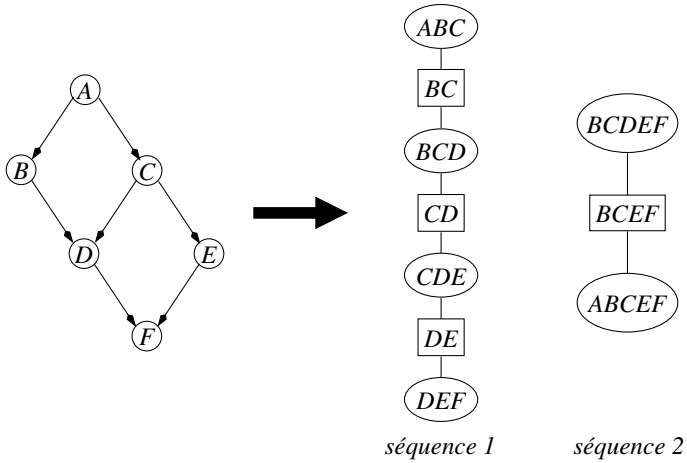
1. S.L. Lauritzen, D.J. Spiegelhalter (1988) *Local computations with probabilities on graphical structures and their application to expert systems(with discussion)*, Journal of the Royal Statistical Society, *Series B*, 50, pp.157–224.
2. F.V. Jensen, S.L. Lauritzen, K.G. Olesen (1990) *Bayesian Updating in Causal Probabilistic Networks by Local Computations*, Comp. Stat. Quarterly, 4, pp.269–282.
3. A.L. Madsen, F.V. Jensen (1998) *Lazy Propagation in Junction Trees*, Proceedings d'UAI-98.
4. A.L. Madsen, F.V. Jensen (1999) *Lazy Propagation : A Junction Tree Inference Algorithm Based on Lazy Evaluation*, Artificial Intelligence, 113, pp.203–245.

❸ Comment construire l'arbre de jonction

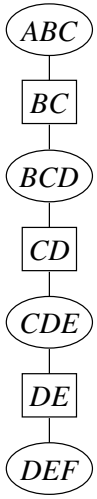
ou la triangulation par l'exemple

Toutes les séquences d'élimination ne sont pas égales

Séquence 1 : A, B, C, F, D, E *Séquence 2 : D, C, A, E, B, F*



De la propriété d'intersection courante



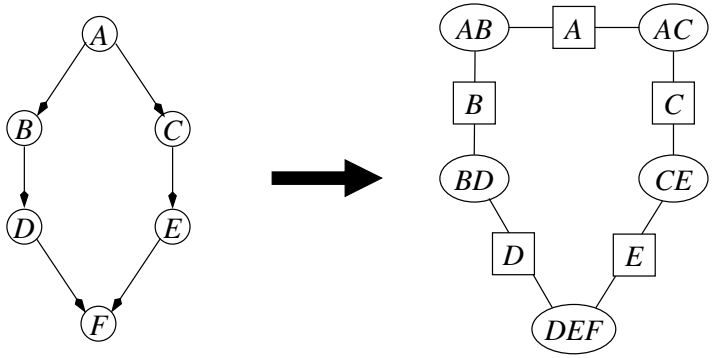
Propriété d'intersection courante

Soient C_1 et C_2 deux cliques quelconques de l'arbre de jonction et soit $S = C_1 \cap C_2 \neq \emptyset$. Alors sur toute chaîne reliant C_1 et C_2, les cliques et séparateurs contiennent S

Théorème

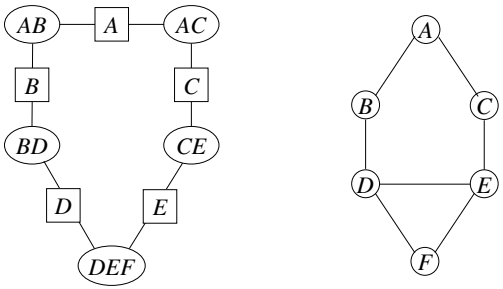
L'algorithme de Shafer-Shenoy fonctionne avec n'importe quel graphe sans cycle vérifiant la propriété d'intersection courante (ce que l'on appelle un *join tree*)

Un graphe de jonction avec cycles



Si l'on n'y prend garde, des cycles peuvent exister bien que la propriété d'intersection courante soit vérifiée

Graphe de jonction et triangulation (1/2)

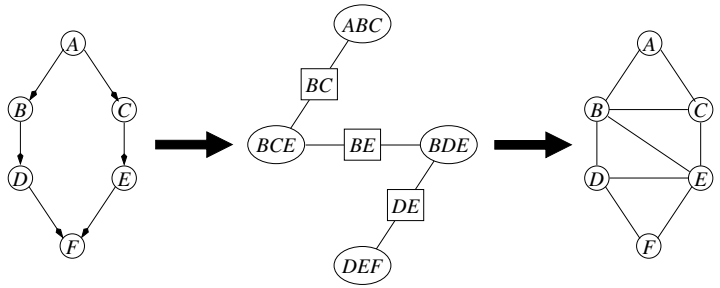


Triangulation

Un graphe non orienté est triangulé si et seulement si, pour tout cycle de longueur 4 ou plus, il existe une corde, c'est-à-dire une arête reliant deux nœuds non consécutifs du cycle

Exemple : le graphe ci-dessus n'est pas triangulé car le cycle A, B, D, E, C, A ne comporte pas de corde

Graphe de jonction et triangulation (2/2)



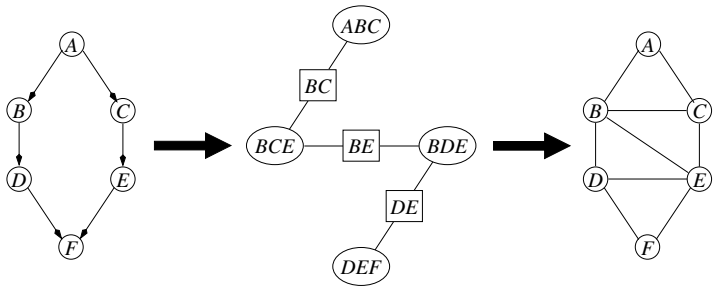
Proposition

il y a équivalence entre les deux assertions :

1. Le graphe de jonction est acyclique
2. Le graphe non orienté correspondant est triangulé

$\Longrightarrow$ pour trouver un « bon » arbre de jonction, il faut trouver une « bonne » triangulation

Un peu de morale, ça ne fait pas de mal



Moralisation

relier tous les parents d'un même nœud, puis supprimer les orientations $\Longrightarrow$ le graphe moral.

$\Longrightarrow$ les cliques pourront contenir l'ensemble des probabilités conditionnelles de la décomposition de la loi jointe

Recherche des triangulations optimales (1/2)

Proposition (Rose 1970)

Un graphe non orienté est triangulé si et seulement si l'application des deux règles suivantes permet d'éliminer tous les nœuds X_i du graphe sans rajouter une seule arête :

1. on rajoute des arêtes entre tous les voisins du nœud X_i que l'on veut éliminer (on forme une clique)
2. on supprime X_i ainsi que les arêtes qui lui sont adjacentes du graphe

$\Longrightarrow$ pour créer un join tree, il suffit de partir d'un graphe non orienté et d'appliquer, avec une certaine séquence d'élimination, les deux points ci-dessus

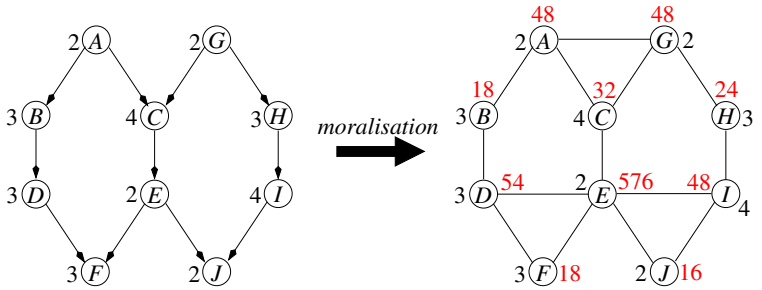
Mellouli (87) : Tout join tree « optimal » peut être construit à partir d'une séquence d'élimination

Recherche des triangulations optimales (2/2)

- *Arnborg et al. (87) :* trouver la triangulation optimale est NP-difficile $\Longrightarrow$ essayer de trouver des heuristiques
- *Kjærulff (90) :* un algorithme glouton rapide et efficace :
  Soit un graphe non orienté (moral) $G = (X, E)$, $X = \{X_1, \ldots, X_n\}$
  1. Associer à chaque X_i un « poids » égal au produit des modalités de X_i et de ses voisins
  2. éliminer le nœud X_i dont le poids est minimal (i.e., relier tous ses voisins de manière à former une clique C_i puis éliminer X_i et ses arêtes adjacentes)
  3. mettre à jour les poids des nœuds restants

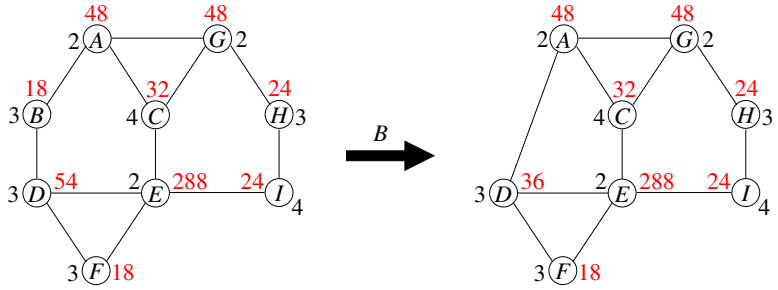
  $\Longrightarrow$ les C_i sont les cliques (ellipses) du join tree
- *van den Eijkhof & Bodlaender (2002) :* "safe reductions"
  $\Longrightarrow$ élimination de variables avec garantie d'optimalité
- Autres algorithmes : Becker & Geiger (96) ; Shoiket & Geiger (87)

Exemple de création de join tree (1/5)



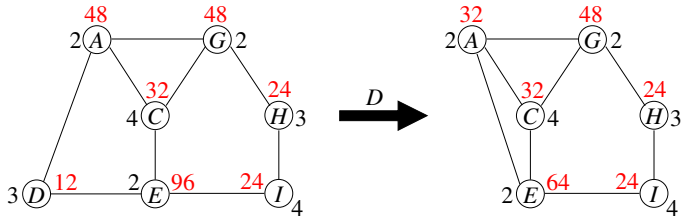
Variable à éliminer : $J \Longrightarrow$ clique EIJ

Exemple de création de join tree (2/5)



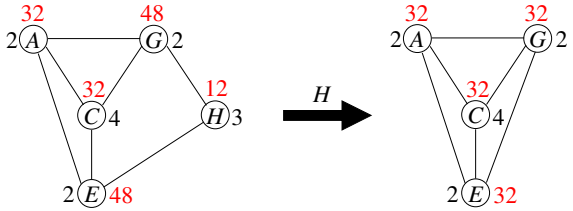
première variable à éliminer : $B \Longrightarrow$ clique ABD
deuxième variable à éliminer : $F \Longrightarrow$ clique DEF

Exemple de création de join tree (3/5)



première variable à éliminer : $D \Longrightarrow$ clique ADE
deuxième variable à éliminer : $I \Longrightarrow$ clique EHI

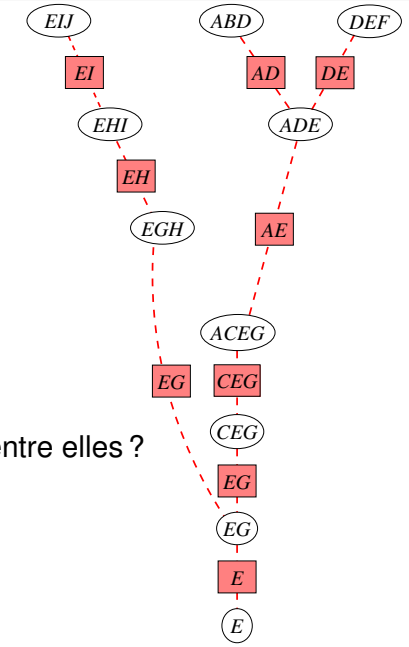
Exemple de création de join tree (4/5)



- première variable à éliminer : $H \Longrightarrow$ clique EGH
- deuxième variable à éliminer : $A \Longrightarrow$ clique $ACEG$
- puis les autres variables peuvent être éliminées dans n'importe quel ordre puisqu'elles appartiennent toutes à la même clique : $C \Longrightarrow$ clique CEG
  $G \Longrightarrow$ clique EG
  $E \Longrightarrow$ clique E

Exemple de création de join tree (5/5)

Ensemble des cliques selon leur ordre de création (avec la variable dont l'élimination a créé la clique) :
EIJ (*J*), *ABD* (*B*), *DEF* (*F*), *ADE* (*D*), *EHI* (*I*), *EGH* (*H*), *ACEG* (*A*), *CEG* (*C*), *EG* (*G*), *E* (*E*)



Problème : comment relier les cliques entre elles ?

Des cliques vers l'arbre d'élimination (1/2)

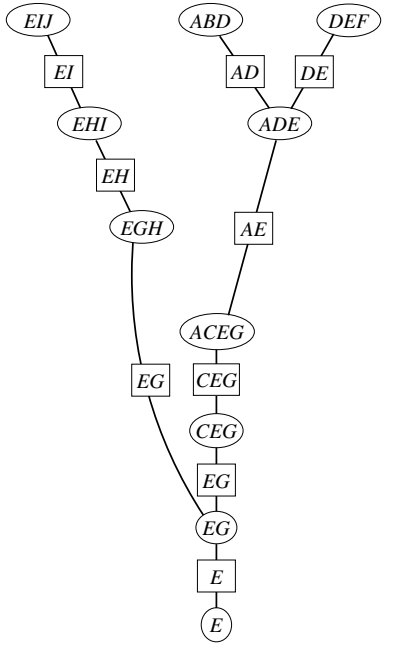
Définition de l'arbre d'élimination

- Soit $\sigma : \{1,\ldots,n\} \mapsto \{1,\ldots,n\}$ la permutation telle que les variables X_i sont éliminées dans l'ordre $X_{\sigma(1)},\ldots,X_{\sigma(n)}$
- Pour tout i, soit $D_{\sigma(i)}$ la clique créée au moment où $X_{\sigma(i)}$ est éliminée
- Arbre d'élimination : graphe $\mathcal{G} = (\mathcal{D}, \mathcal{E})$, où :
  - $\mathcal{D} = \{D_{\sigma(i)} : i \in \{1,\ldots,n\}\}$,
  - $\mathcal{E} = \{(D_{\sigma(i)}, D_{\sigma(j)}) :$
    $1 \leq i < n, j = min\{k \neq i : X_{\sigma(k)} \in D_{\sigma(i)}\}\}$

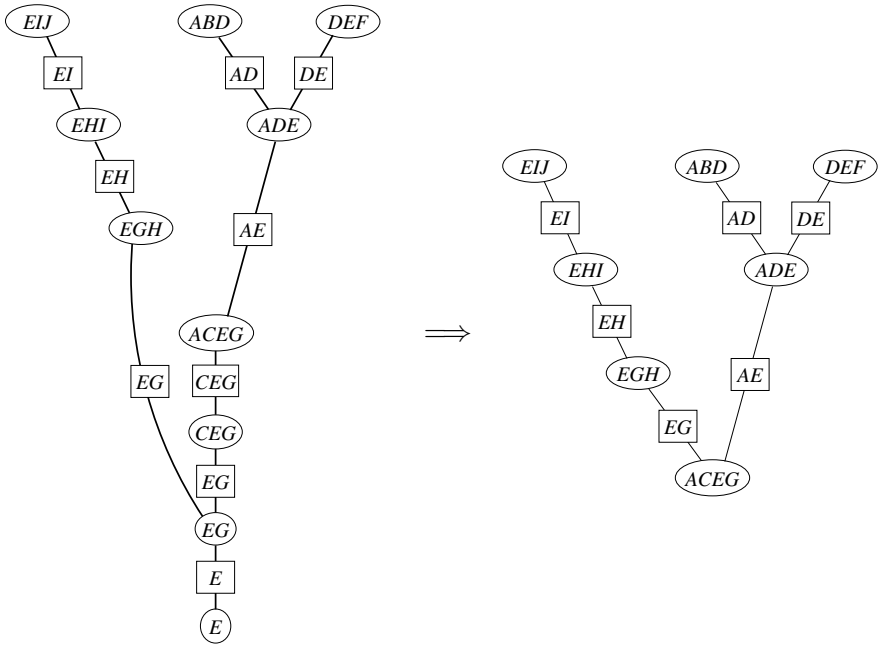
$\Longrightarrow$ Si l'on trie les nœuds X_i à l'intérieur des cliques selon leur ordre d'élimination, alors :
on relie $D_{\sigma(i)} = \{X_{\sigma(i)}, X_{\sigma(j)}, \ldots\}$ à $D_{\sigma(j)}$.

Des cliques vers l'arbre d'élimination (2/2)

- Arbre d'élimination : $\mathcal{G} = (\mathcal{D}, \mathcal{E})$, où :
  - $\mathcal{D} = \{D_{\sigma(i)} : i \in \{1,\ldots,n\}\}$,
  - $\mathcal{E} = \{(D_{\sigma(i)}, D_{\sigma(j)}) : 1 \leq i < n,$
    $j = min\{k \neq i : X_{\sigma(k)} \in D_{\sigma(i)}\}\}$
- Ensemble des cliques selon leur ordre de création (avec la variable dont l'élimination a créé la clique) :
  EIJ (*J*), *ABD* (*B*), *DEF* (*F*), *ADE* (*D*), *EHI* (*I*), *EGH* (*H*), *ACEG* (*A*), *CEG* (*C*), *EG* (*G*), *E* (*E*)



De l'arbre d'élimination vers l'arbre de jonction (1/2)



Propriétés de l'arbre d'élimination

Propriétés

$\mathcal{D} = \{D_{\sigma(i)} : i \in \{1, \ldots, n\}\}$,
$\mathcal{E} = \{(D_{\sigma(i)}, D_{\sigma(j)}) : 1 \leq i < n, j = min\{k \neq i : X_{\sigma(k)} \in D_{\sigma(i)}\}\}$

1. L'arbre d'élimination est un arbre
2. Il vérifie la propriété d'intersection courante
3. Soit $D_{\sigma(j)}$ un enfant de $D_{\sigma(i)}$, alors $|D_{\sigma(j)}| \geq |D_{\sigma(i)}| - 1$
4. Soient $D_{\sigma(i)}$ et $D_{\sigma(j)}$ les parents de $D_{\sigma(k)}$, alors $D_{\sigma(i)} \not\subset D_{\sigma(j)}$ et $D_{\sigma(j)} \not\subset D_{\sigma(i)}$
5. Soit $D_{\sigma(j)}$ un enfant de $D_{\sigma(i)}$, alors $D_{\sigma(j)} \subset D_{\sigma(i)} \iff |D_{\sigma(j)}| = |D_{\sigma(i)}| - 1$
6. Soit $D_{\sigma(j)}$ un enfant de $D_{\sigma(i)}$ tel que $D_{\sigma(j)} \not\subset D_{\sigma(i)}$, alors il n'existe pas d'ancêtre $D_{\sigma(k)}$ de $D_{\sigma(i)}$ tel que $D_{\sigma(j)} \subset D_{\sigma(k)}$

De l'arbre d'élimination vers l'arbre de jonction (2/2)

Algorithme pour obtenir un arbre de jonction

01 créer l'arbre d'élimination $\mathcal{G} = (\mathcal{D}, \mathcal{E})$
02 marquer à **false** tous les arcs de $\mathcal{E}$
03 **pour** i variant de n à 1 **faire**
04 **si** il existe $D_{\sigma(j)}$ parent de $D_{\sigma(i)}$ tel que l'arc $(D_{\sigma(j)}, D_{\sigma(i)})$ est non marqué et $|D_{\sigma(i)}| = |D_{\sigma(j)}| - 1$ **alors**
05 **pour** tous les autres parents $D_{\sigma(k)}$ de $D_{\sigma(i)}$ **faire**
06 créer dans $\mathcal{G}$ un arc $(D_{\sigma(k)}, D_{\sigma(j)})$
07 marquer cet arc à `true`
08 **fait**
09 **si** $D_{\sigma(i)}$ a un enfant $D_{\sigma(k)}$ **alors**
10 créer dans $\mathcal{G}$ un arc $(D_{\sigma(j)}, D_{\sigma(k)})$
11 **finsi**
12 supprimer $D_{\sigma(i)}$ ainsi que ses arcs adjacents
13 **fait**

A la fin de l'algorithme ci-dessus, $\mathcal{G}$ est un arbre de jonction.

Améliorations de l'algorithme d'élimination

- Au lieu de choisir le nœud à éliminer en fonction du poids, choisir, quand c'est possible, un nœud appartenant à une seule clique (« simplicial rule », bodlaender (02)) ;
- Sous certaines contraintes, choisir un nœud presque simplicial (il manque une seule arête pour former une clique) ;
- Autres règles de réduction optimales (Buddies rule, Extended cube rule...) ;
- Suppression des arêtes de triangulation superflues (cf. Kjærulff (90)) — nécessite de recalculer la séquence d'élimination (par exemple par maximum cardinality search) ;
- Optimisation de l'arbre de jonction par modification des adjacences (cf. Jensen & Jensen (94)).

Références bibliographiques (1/2)

Quelques références

- Kjærulff, U (1990) *Triangulation of graphs – Algorithms giving small total state space*, technical report.
- Kjærulff, U (1991) *Optimal decomposition of probabilistic networks by simulated annealing*, Statistics and Computing, Vol 2, pp7-17.
- Becker, A & Geiger, D (1996) *A sufficiently fast algorithm for finding close to optimal junction trees*, Proceedings d'UAI-96.
- van den Eijkhof, F & Bodlaender A (2002) *Safe reduction rules for weighted treewidth*, Proceedings of the 28th International Workshop on Graph-Theoretic Concepts in Computer Science, Lecture Notes in Computer Science, vol 2573, pp176–185.

Références bibliographiques (2/2)

Quelques références

- Jensen, F.V. & Jensen, F. (1994) *Optimal junction trees* Proceedings d'UAI-94.
- Shoikhet, K & Geiger, D (1997) *Finding optimal triangulations via minimal vertex separators*, Proccedings de AAAI-97.
- Leimer, H.-G. (1993) *Optimal decomposition by clique separators*, Discrete Mathematics, vol 113, pp99-123.
- Olesen, K & Madsen, A (1999) *Maximal prime decomposition of Bayesian networks*, technical report.
- Flores, J & Gámez, J & Olesen, K (2003) *Incremental compilation of Bayesian networks*, Proceedings d'UAI-03.