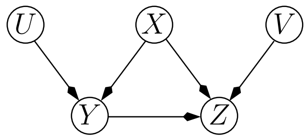
Examen du module MGDE

Durée : 3 heures

Exercice 1 (5 points)

La loi jointe $p(u, x, v, y, z)$ des 5 v.a. (U, X, V, Y, Z) admet le graphe d'indépendance G de la figure ci-dessous :



Q 1.1 Qu'est-ce que le critère de d-séparation permet d'affirmer concernant les propriétés suivantes :

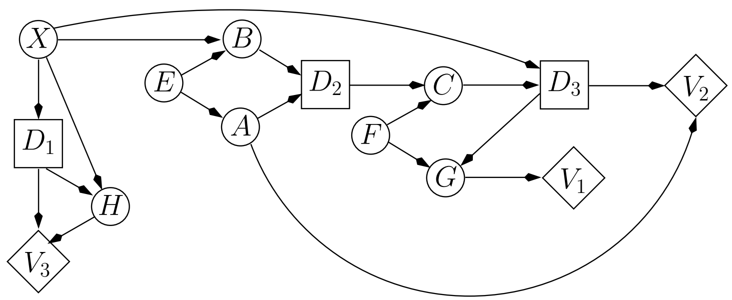
$V \perp\!\!\!\perp (X, Y)$? $U \perp\!\!\!\perp (V, X)$? $U \perp\!\!\!\perp V \mid (Y, Z)$?
$U \perp\!\!\!\perp V \mid (X, Y)$? $U \perp\!\!\!\perp V \mid (X, Y, Z)$?

Q 1.2 L'ordre d'énumération sur les variables utilisé a été $UXVYZ$; quelles sont les relations d'indépendance et d'indépendance conditionnelle qui ont servi à construire le graphe ?
Comment $p(u, x, v, y, z)$ se décompose-t-elle ?

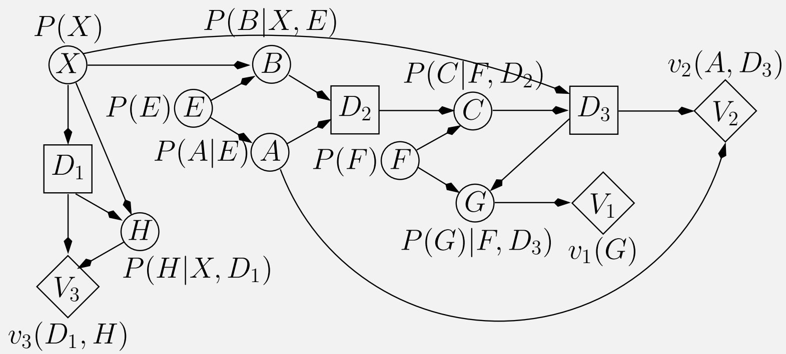
Q 1.3 À partir des relations explicitées à la question 2 et des propriétés générales de la relation $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ démontrez la validité des assertions qui avaient été obtenues à la question 1 par d-séparation ainsi que celle de $U \perp\!\!\!\perp (V,Z) \mid (X,Y)$ et de $U \perp\!\!\!\perp Z \mid (X,V,Y)$.

Exercice 2 (4,5 points)

Soit le diagramme d'influence ci-dessous :



Q 2.1 Indiquez à côté des différents nœuds de ce graphe quelles tables vous stockeriez dans ceux-ci (par exemple, des tables du type $P(A|B)$, $u(X,Y)$, etc).



Q 2.2 Déterminez un ordre partiel temporel des variables, puis un ordre total compatible, sachant que les décisions sont prises dans l'ordre D_1, D_2, D_3.

Pour déterminer un ordre partiel, il faut déterminer les variables aléatoires dont la valeur est connue avec certitude avant de prendre les décisions D_1, D_2, D_3. Avant D_1, seul X est connu. Avant D_2, A et B sont connus. Avant D_3, C est aussi connu. D'où l'ordre partiel :

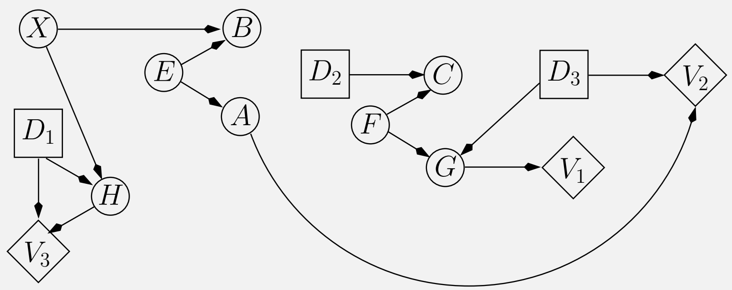
$$\{X\} \prec D_1 \prec \{A, B\} \prec D_2 \prec \{C\} \prec D_3 \prec \{E, H, F, G\}.$$

Ordre total compatible :

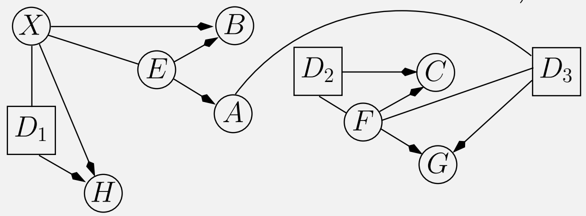
$$X \prec D_1 \prec A \prec B \prec D_2 \prec C \prec D_3 \prec E \prec H \prec F \prec G. \tag{1}$$

Q 2.3 En utilisant l'ordre total précédent, créez un « strong junction tree ». Vous préciserez dans quelles cliques vous stockerez les tables indiquées à la question 1.

Le réseau de valuation est :



Après moralisation et suppression des nœuds d'utilité, il devient :



L'élimination des nœuds tels qu'indiqué par l'équation (1) nous donne alors les cliques et sous-cliques suivantes : FGD_3, CFD_2D_3, HXD_1, $ABEX$, ACD_2D_3, ACD_2, AD_2, ABX, AX, XD_1 et X. Après élimination des sous-cliques et en appliquant la règle du strong junction tree, on obtient l'arbre de jonction suivant :

$$HXD_1 \leftarrow ABEX \leftarrow ACD_2D_3 \leftarrow CFD_2D_3 \leftarrow FGD_3$$

On peut stocker les tables de la manière suivante :

clique	tables
HXD_1	$P(X), P(H\|X, D_1), v_3(D_1, H)$
$ABEX$	$P(E), P(A\|E), P(B\|X, E)$
ACD_2D_3	$v_2(A, D_3)$
CFD_2D_3	$P(F), P(C\|F, D_2)$
FGD_3	$P(G\|F, D_3), v_1(G)$

Exercice 3 (5 points)

Une entreprise fabrique des puces électroniques et cherche à améliorer l'efficacité (variable E) de ses ouvriers. L'efficacité d'un(-e) ouvrier(-ère) est liée à sa capacité de concentration (variable C), sa formation technique (variable F) et sa dextérité manuelle (variable D). Le chef d'entreprise dispose d'une base de données, constituée d'extraits des fiches personnelles de chacun de ses employés passés ou actuels, lui permettant d'estimer la loi jointe $p(c, f, d, e)$ de ces 4 variables. En revanche, la loi lui interdisant toute discrimination selon l'âge (variable A) ou le sexe (variable S), ces deux variables ne sont pas retenues dans la base, bien qu'il sache que :

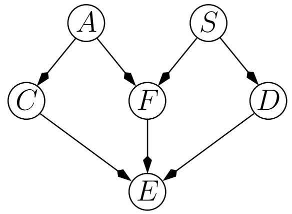
- degré de concentration et niveau de formation dépendent de l'âge ;
- degré de dextérité et niveau de formation dépendent du sexe.

Q 3.1 L'analyse de la loi jointe des v.a. (C, F, D, E) montre que :

- Ces v.a. sont deux à deux dépendantes dans tout contexte sauf la paire (C, D) ;
- $C \perp\!\!\!\perp D \mid F$; NON$[C \perp\!\!\!\perp D \mid E]$; NON$[C \perp\!\!\!\perp D \mid (F, E)]$.

À quel graphe marqué l'application de l'algorithme IC^* aboutit-elle ?

Q 3.2 On admet la validité du graphe causal de la figure ci-dessous.



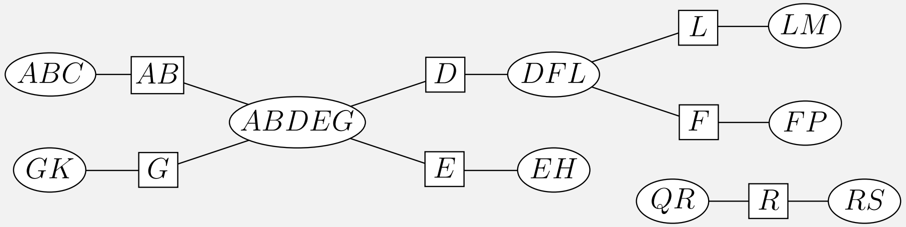
Le chef d'entreprise se demande dans quelle mesure une meilleure formation améliorerait l'efficacité de ses ouvriers ; pour cela, il veut évaluer l'effet $p(e \mid\mid f^*)$ sur E d'une intervention sur F.
Y a-t-il une ou des variable(s) permettant d'identifier cet effet par blocage en amont ?
Si oui, quelle expression obtient-on pour $p(e \mid\mid f^*)$?

Exercice 4 (6,5 points)

Considérons une relation de préférence $\succsim$ sur un ensemble $\mathcal{X} = A \times B \times C \times D \times E \times F \times G \times H \times K \times L \times M \times P \times Q \times R \times S$ représentable par la fonction d'utilité GAI-décomposable :

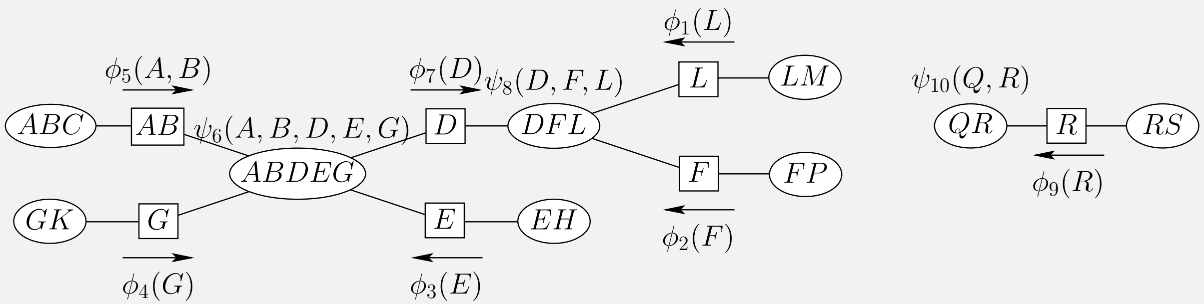
$$u_1(A, B, D, E, G)+u_2(A, B, C)+u_3(D, F, L)+u_4(E, H)+u_5(G, K)+u_6(F, P)+u_7(L, M)+u_8(Q$$

Q 4.1 Dessinez le réseau GAI représentant la fonction d'utilité ci-dessus.



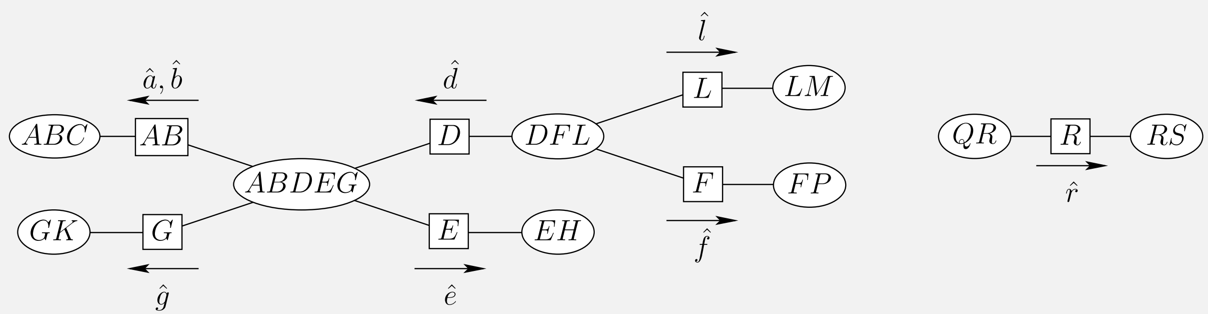
Q 4.2 Supposons que l'on veuille calculer l'élément préféré selon $\succsim$ dans $\mathcal{X}$. Indiquez sur le réseau GAI quels messages vous feriez transiter lors d'une phase de collecte vers les cliques DFL et QR pour effectuer ce calcul.

Les messages ci-dessous sont calculés avec la séquence d'élimination de cliques suivante : LM, FP, EH, GK, ABC, $ABDEG$, RS, et ils sont numérotés dans l'ordre dans lesquels ils sont calculés.



$$
\begin{aligned}
\phi_1(L) &= \max_M u_7(L,M)\\
\phi_2(F) &= \max_P u_6(F,P)\\
\phi_3(E) &= \max_H u_4(E,H)\\
\phi_4(G) &= \max_K u_5(G,K)\\
\phi_5(A,B) &= \max_C u_2(A,B,C)\\
\psi_6(A,B,D,E,G) &= = u_1(A,B,D,E,G) + \phi_3(E) + \phi_4(G) + \phi_5(A,B)\\
\phi_7(D) &= \max_{A,B,E,G} \psi_6(A,B,D,E,G)\\
\psi_8(D,F,L) &= u_3(D,F,L) + \phi_1(L) + \phi_2(F) + \phi_7(D)\\
\phi_9(R) &= \max_S u_9(R,S)\\
\psi_{10}(Q,R) &= u_8(Q,R) + \phi_9(R)
\end{aligned}
$$

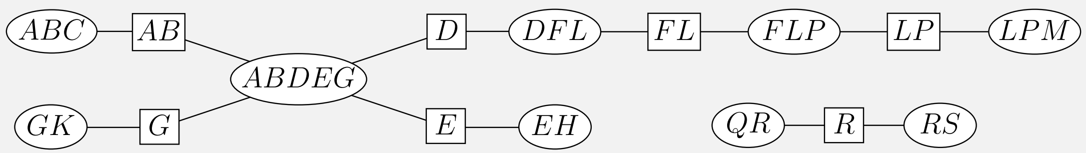
Q 4.3 En notant $\hat{a}$, $\hat{b}$, etc, les valeurs des attributs à l'optimum, indiquez les messages propagés à partir des cliques DFL et QR lors de la phase de diffusion.



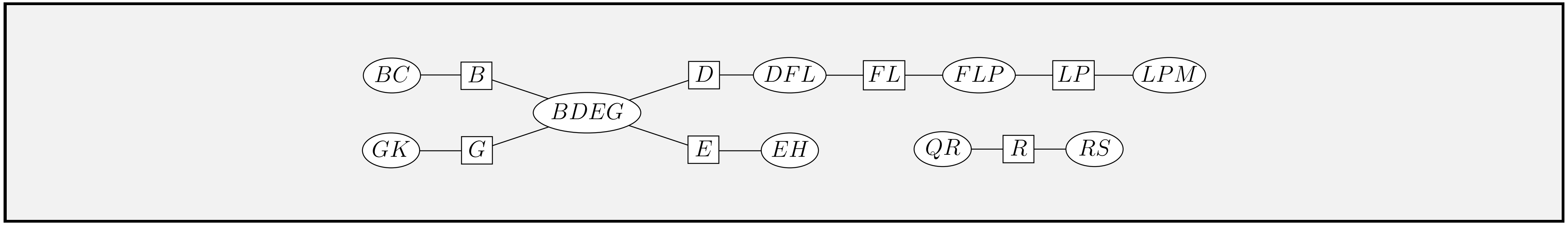
$$
\begin{aligned}
(\hat{q}, \hat{r}) &= \text{Argmax}\{\psi_{10}(Q, R)\}\\
(\hat{d}, \hat{f}, \hat{l}) &= \text{Argmax}\{\psi_8(D, F, L)\}\\
(\hat{a}, \hat{b}, \hat{d}, \hat{e}, \hat{g}) &= \text{Argmax}\{\psi_6(A, B, \hat{d}, E, G)\}\\
(\hat{a}, \hat{b}, \hat{c}) &= \text{Argmax}\{u_2(\hat{a}, \hat{b}, C)\}\\
(\hat{g}, \hat{k}) &= \text{Argmax}\{u_5(\hat{g}, K)\}\\
(\hat{e}, \hat{h}) &= \text{Argmax}\{u_4(\hat{e}, H)\}
\end{aligned}
$$

Q 4.4 Si l'on rajoute la contrainte $M \neq P$, quel nouveau réseau GAI obtient-on ?

Rajouter une contrainte entre M et P revient à rajouter une utilité $u_{10}(M, P)$, ce qui nécessite une retriangulation du réseau markovien, qui demande l'ajout soit d'une arête (L, P) ou (F, M). Si l'on rajoute l'arête (L, P), on obtient le réseau :

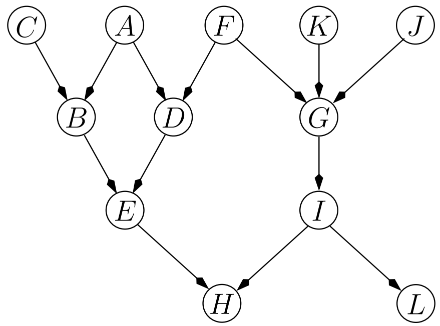


Q 4.5 Si l'on rajoute en outre la contrainte $A = a_0$, c'est-à-dire que l'on fixe la valeur de l'attribut A à a_0, quel nouveau réseau GAI obtient-on ?

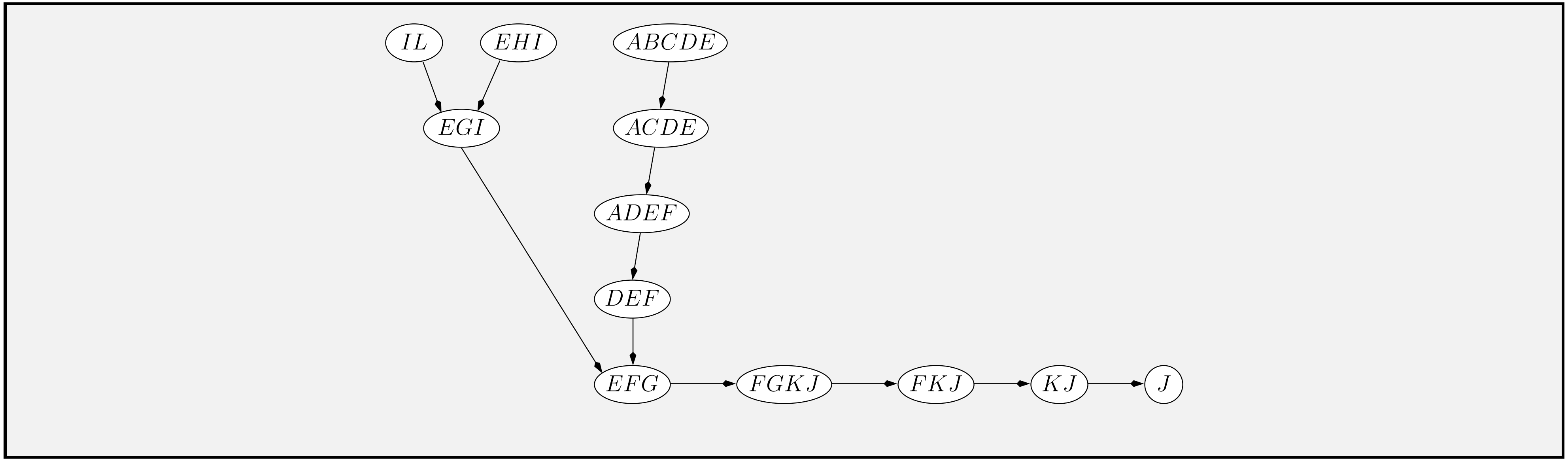


Exercice 5 (4 points)

Considérons le réseau bayésien suivant :



Q 5.1 Dessinez l'arbre d'élimination correspondant à la séquence d'élimination de variables : $L, B, C, A, H, D, I, E, G, F, K, J$.



Q 5.2 Dessinez l'arbre de jonction obtenu à partir de l'arbre d'élimination en utilisant l'algorithme vu en cours.