

USTHB

Faculté d'Informatique

Département Intelligence Artificielle et Sciences des Données

Master 2 Informatique visuelle

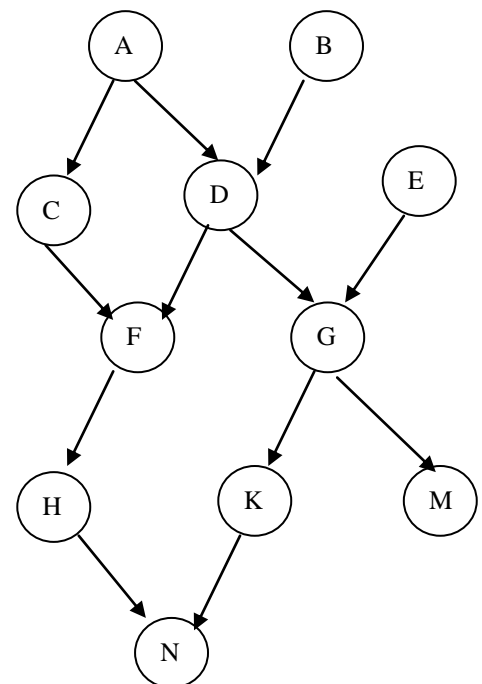
Représentation des Connaissances et Raisonnement

Année Universitaire : 2023-2024

TD N° 6-TP N°6
Réseaux Bayésiens

Exercice 1 :

Construisez l'arbre de jonction associé au graphe
à connexions multiples suivants :



Principalement, l'algorithme d'inférence exact pour les graphes à connexions multiples se comporte de la manière suivante :

- **La phase de construction :** Elle nécessite un ensemble de sous-étapes afin de transformer le graphe initial en un arbre de jonction, dont les nœuds sont des **clusters** (regroupements des nœuds du graphe initial). Cette transformation est nécessaire pour :
 - Éliminer les boucles du graphe initial,
 - Optimiser l'algorithme d'inférence.

Elle s'effectue en trois étapes :

1. Moralisation du DAG initial \mathcal{G} ,
 2. Triangulation du graphe moral,
 3. Création de l'arbre de jonction à partir du graphe moral triangulé.
- **La phase de propagation :** C'est la phase de calcul de l'ensemble des distributions de probabilités du réseau après l'avènement de nouvelles informations. Cette phase s'effectue par le biais de passage de messages entre les nœuds de l'arbre de jonction.

Étape 1 : Moralisation : Elle consiste à marier deux à deux les parents de chaque variable en les reliant par un arc non-dirigé. À partir d'un graphe \mathcal{G} , le graphe moral associé noté \mathcal{G}_M est obtenu par la procédure suivante :

Algorithme 3.2 : Construction du graphe moral [12, 82]

Début

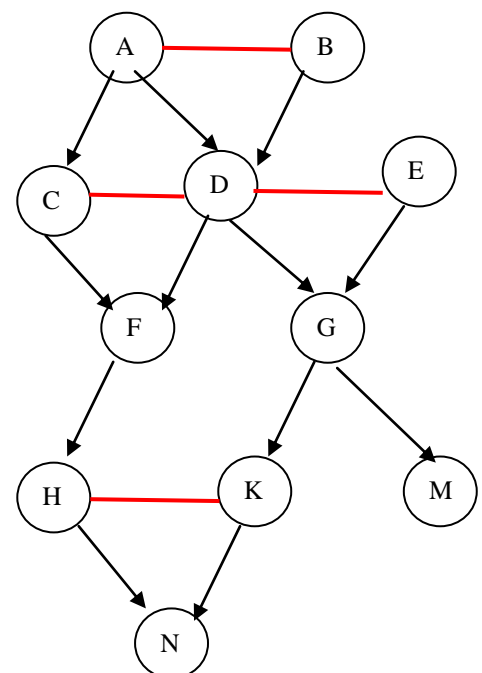
Associer au DAG initial \mathcal{G} un graphe non dirigé en éliminant les directions de tous les arcs,

Construire le graphe moral \mathcal{G}_M à partir de \mathcal{G} en reliant les parents de chaque nœud en rajoutant des arcs non dirigés.

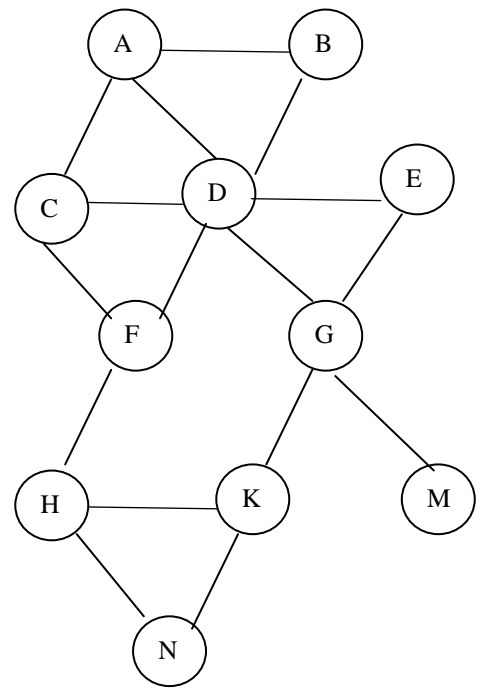
Fin

Etape1 : La moralisation :

La moralisation a permis l'ajout de 4 arcs non orientés



Graphe moralisé obtenu



Étape 2 : Triangulation du graphe moral :

En général, il existe plusieurs façons de trianguler un graphe moral. La tâche de trouver la triangulation la plus optimale est NP-complet [83, 84, 78]. Néanmoins, plusieurs heuristiques ont été développées afin de réduire le coût induit par l'inférence, parmi lesquelles se trouve le critère de sélection de nœuds qui est une heuristique (induisant une complexité polynômial) produisant une triangulation de haute qualité [85].

La procédure de triangulation qui suit permet également d'identifier l'ensemble des clusters ou cliques (noté *cluster-set*) en utilisant l'algorithme de Golumbic [86] qui garantit qu'aucun cluster n'est inclus dans un autre cluster plus grand.

Algorithme 3.3 : Triangulation du graphe moral et identification des clus-

ters [85]

Debut

Cluster-set := \emptyset ;

$\mathcal{G}'_{\mathcal{M}} := \mathcal{G}_{\mathcal{M}}$;

Tantque il existe des nœuds dans $\mathcal{G}'_{\mathcal{M}}$ Faire

Si il existe un nœud A tel que tous ses nœuds adjacents sont connectés

Alors

Créer un cluster C contenant A ainsi que ses nœuds adjacents ;

Supprimer A de $\mathcal{G}'_{\mathcal{M}}$

Sinon

Sélectionner un nœud A ayant le nombre de nœuds adjacents le plus petit ;

Ajouter des arcs afin de connecter l'ensemble des nœuds adjacents ;

Pour chaque arc ajouté à $\mathcal{G}'_{\mathcal{M}}$, ajouter l'arc correspondant à $\mathcal{G}_{\mathcal{M}}$;

Créer un cluster C contenant A ainsi que ses nœuds adjacents ;

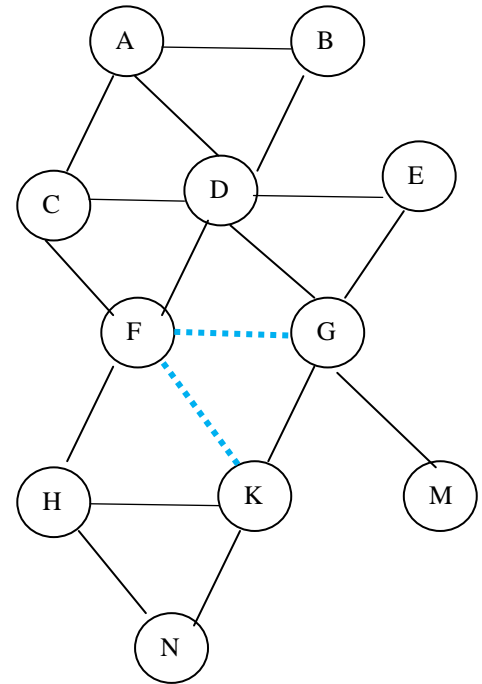
Supprimer A de $\mathcal{G}'_{\mathcal{M}}$

Si $C \notin \text{cluster-set}$ Alors $\text{cluster-set} := \text{cluster-set} \cup \{C\}$;

Fin

Etape 2 : La triangulation :

Graphe triangulé



Etape 3 : Construction de l'arbre de jonction :

Debut

Pour $i := 1$ à m *Faire*

Separator-set $:= \emptyset$;

$C_i := \text{Cluster-set}[i]$;

Pour $j := 1$ à $(m-1)$ *Faire*

$C_j := \text{Cluster-set}[j]$;

Créer un séparateur candidat S_{ij} pour les deux clusters C_i et C_j ;

Insérer S_{ij} dans *Separator-set*;

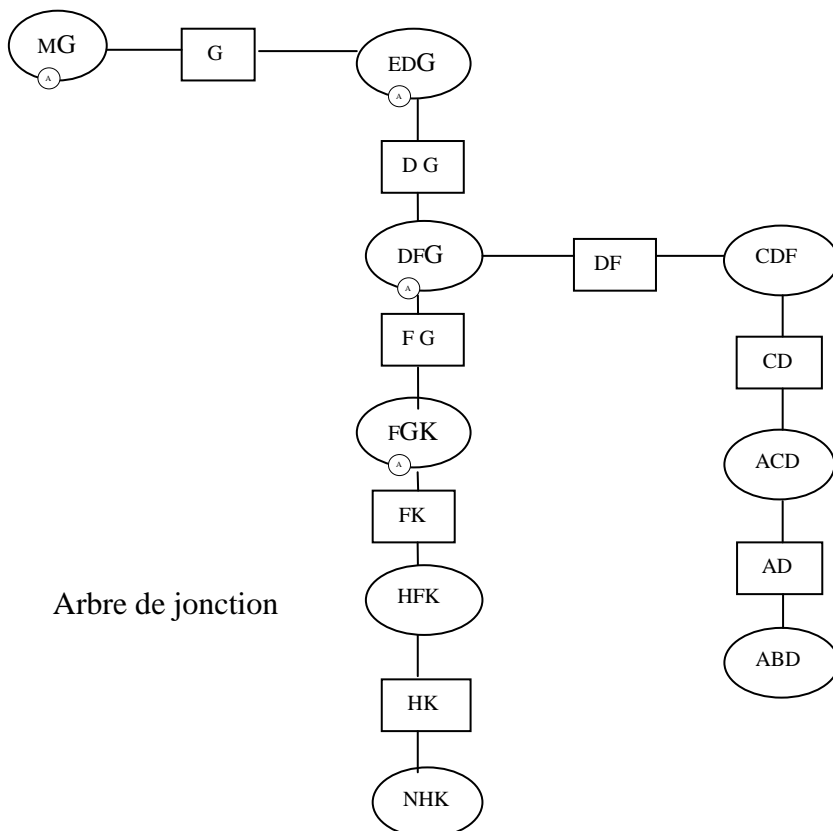
Sélectionner un séparateur S_{ij} à partir de *Separator-set*, en accord avec le critère de sélection spécifié ci-après ;

Insérer le séparateur S_{ij} entre le cluster C_i et le cluster C_j

Fin

Nœud éliminé	Cluster induit	Arcs ajoutés	Cluster set
M	MG	rien	{MG}
B	ABD	rien	{MG,ABD}
E	EDG	rien	{MG,ABD,EDG}
N	NHK	rien	{MG,ABD,EDG,NHK}
A	ACD	rien	{MG,ABD,EDG,NHK,ACD}
C	CDF	rien	{MG,ABD,EDG,NHK,ACD,CDF}
D	DFG	FG	{MG,ABD,EDG,NHK,ACD,CDF,DFG}
H	HFK	HK	{MG,ABD,EDG,NHK,ACD,CDF,DFG,HFK}
G	FGK	rien	{MG,ABD,EDG,NHK,ACD,CDF,DFG,HFK,FGK}
F	FK	rien	{MG,ABD,EDG,NHK,ACD,CDF,DFG,HFK,FGK} {FK} \subset {FGK}
K	K	rien	{MG,ABD,EDG,NHK,ACD,CDF,DFG,HFK,FGK} {K} \subset {FGK}

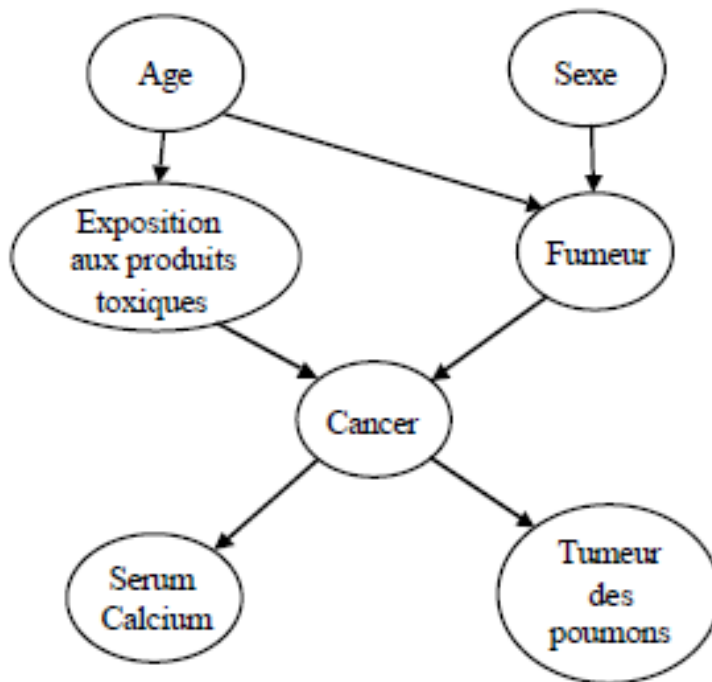
Clusters	Séparateurs candidats	Masse	Coûts	Séparateurs sélectionnés
MG ABD	G	0	0	*
EDG		1	2	
NHK		0	0	
ACD		0	0	
CDF	G	0	0	
DFG		1	2	
HFK		0	0	
FGK	G	1	2	
ABD EDG	D	1	2	*
NHK	AD	0	0	
ACD		2	4	
CDF		1	2	
DFG		1	2	
HFK	D	0	0	
FGK		0	0	
EDG NHK	D	0	0	*
ACD		0	0	
CDF		1	2	
DFG		2	4	
HFK	DG	0	0	
FGK	G	1	2	
NHK ACD	HK	0	0	*
CDF		0	0	
DFG		0	0	
HFK		2	4	
FGK		1	2	
ACD CDF	CD	2	4	*
DFG	D	1	2	
HFK		0	0	
FGK		0	0	
CDF DFG	DF	2	4	*
HFK	F	1	2	
FGK	F	1	2	
DFG HFK	F	1	2	*
FGK	FG	2	4	
HFK FGK	FK	2	4	*



Exercice 2 :

Considérez le problème de conception d'un réseau Bayésien pour le diagnostic du cancer des poumons. Les paramètres à considérer sont : l'âge du patient, le sexe du patient l'exposition à des produits toxiques, le tabagisme, le cancer, la tumeur aux poumons et les carences en calcium.

a- Spécifier la structure du réseau Bayésien



b- Proposez des tables de distributions conditionnelles.

- Si $(U_A) = \emptyset$ (A est un nœud racine) alors, il s'agit de spécifier les probabilités a priori relatives aux différentes instances de la variable A, tout en respectant la condition de normalisation qui stipule que :

$$\sum_a P(a) = 1$$

- Si $(U_A) \neq \emptyset$ alors il faudrait spécifier les probabilités conditionnelles des différentes instances a de A dans le contexte des différentes instances de ses parents u_A telles que :

$$\sum_a P(a \mid u_A) = 1$$

c- Comment se fait le calcul des probabilités jointes.

$$\begin{aligned}
 P(A, S, E, F, C, T, SC) = & \\
 & P(A) \cdot P(S) \cdot \\
 & P(E | A) \cdot P(F | A, S) \cdot \\
 & P(C | E, F) \cdot \\
 & P(SC | C) \cdot P(T | C)
 \end{aligned}$$

Exemple :

$$P(a \wedge \neg s \wedge e \wedge f \wedge \neg c \wedge l \wedge t) = P(a) \cdot P(\neg s) \cdot P(e|a) \cdot P(f|\neg s) \cdot P(\neg c | e \wedge f) \cdot P(l|\neg c) \cdot P(t|\neg c)$$

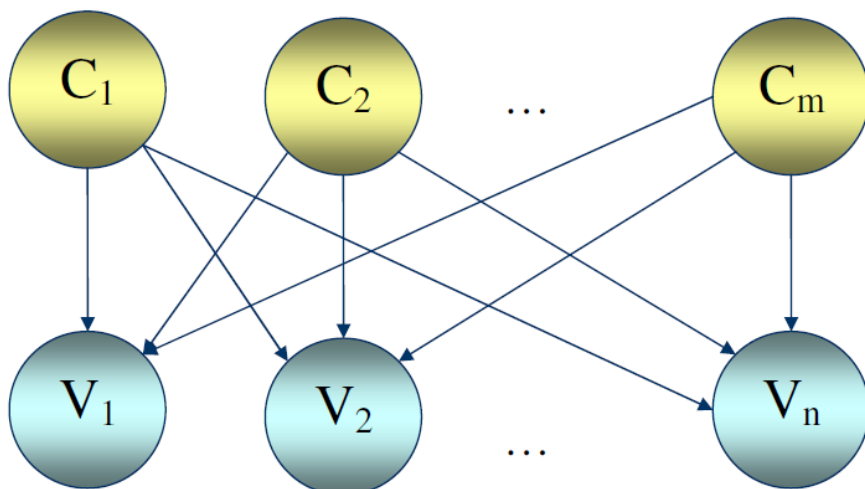
Exercice 3 :

Etant donné un ensemble de N document de textes, le problème est de classer ces documents selon un nombre de classes C_1, \dots, C_m .

Proposez une structure du réseau Bayésien.

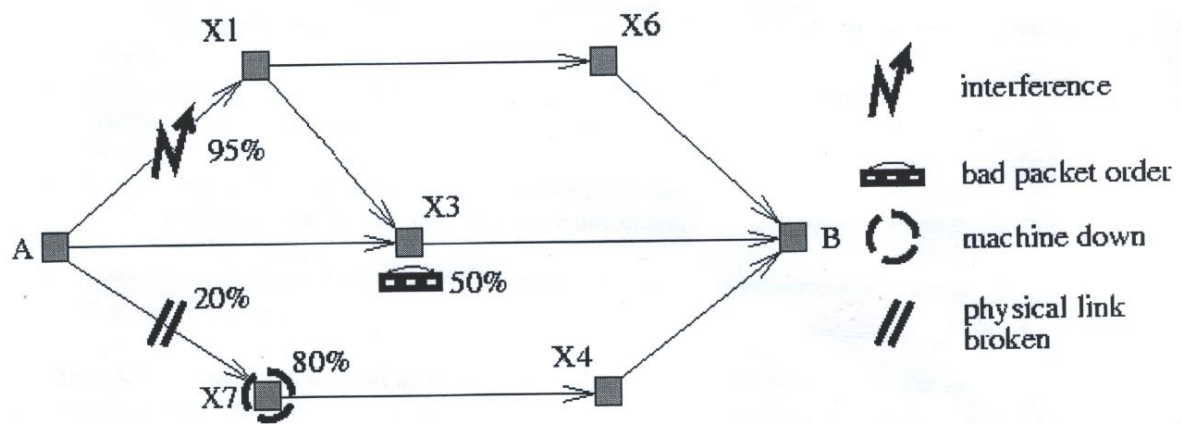
Chaque document est représenté par un vecteur de caractéristiques $X(x_1, \dots, x_m)$. Les caractéristiques sont les termes qui peuvent bien identifier les classes de documents.

Le réseau bayésien utilisé dans ce système est un réseau très simple qui contient de 2 couches, il se base sur une hypothèse que les caractéristiques de documents sont indépendantes. Dans ce réseau les nœuds V_i représentent les caractéristiques et les nœuds C_j représentent les classes :

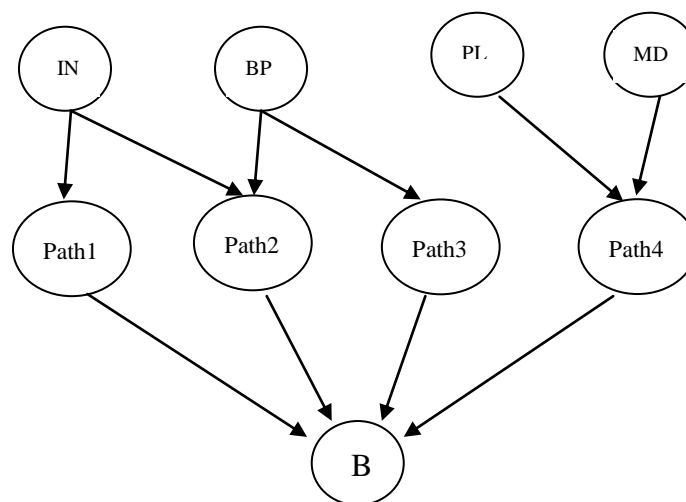


Exercice 4 :

Considérons un réseau de défaillance représenté par la figure suivante. Le nœud A envoie des paquets au nœud B. Seul A émet des paquets.



Modélisez le problème du diagnostic à l'aide d'un réseau Bayésien :



TP N°5

BNT (Bayes Net Toolbox for Matlab)

<https://code.google.com/p/bnt/>

Il s'agit d'exploiter la toolbox BNT sur des polyarbres et des graphes à connexions multiples en :

- Définissant les structures graphiques
- Générant la composante graphique en termes de :
 - probabilités a priori pour les nœuds racine,
 - probabilités conditionnelles,
- Simulant le processus de propagation afin de calculer :
 - $P(\text{variable d'intérêt} \mid \text{évidence(s)})$
- Générant des graphes Générez un graphe à connexions multiples tels que le nombre de variables et le nombre de parents max sont très grands. La construction de l'arbre de jonction étant un problème NP-Complet, la construction de l'arbre associé à un réseau bayésien de très grandes instances peut devenir impossible.