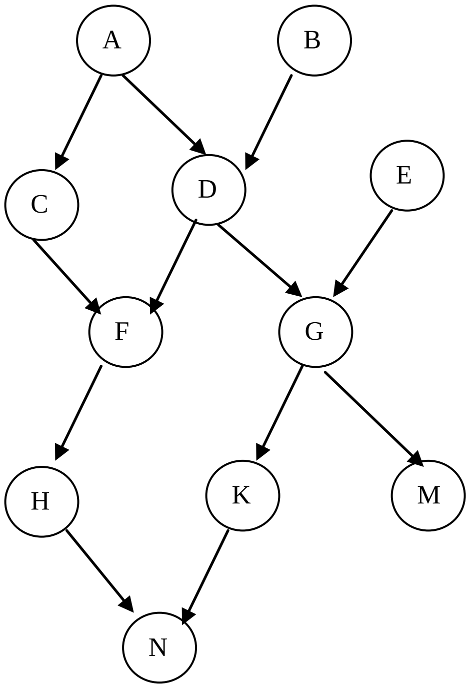
USTHB
Faculté d'Informatique
Département d'Informatique
Master 2 SII — Année Universitaire : 2022-2023
Représentation des connaissances et Raisonnement 2

**TD N° 5 – TP N°4**
Réseaux Bayésiens et Réseaux possibilistes

**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. A partir d'un graphe $\mathcal{G}$, le graphe moral associé noté $\mathcal{G}_{\mathcal{M}}$ est obtenu par la procédure suivante :

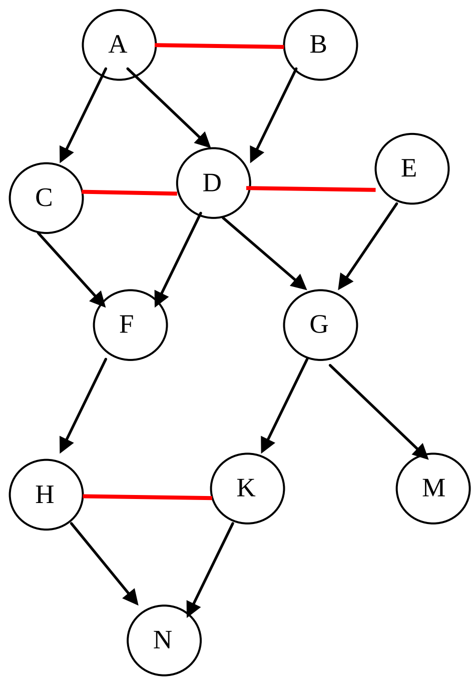
***Algorithme 3.2 : Construction du graphe moral [12, 82]***
***Début***

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

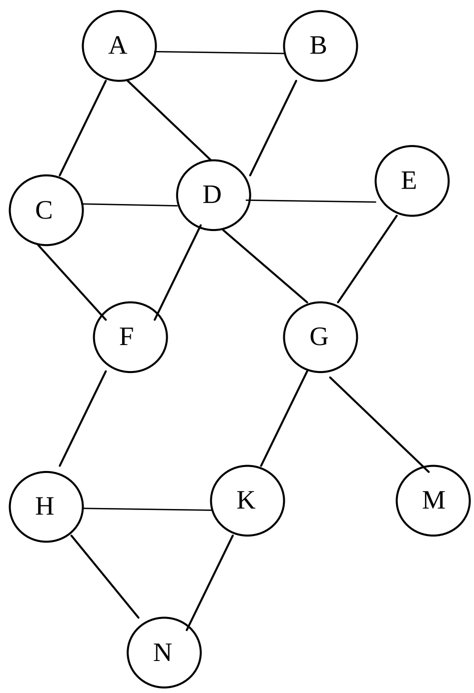
*Construire le graphe moral $\mathcal{G}_{\mathcal{M}}$ à partir de $\mathcal{G}$ en reliant les parents de chaque nœud en rajoutant des arcs non dirigés.*

***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 clusters [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 ;*

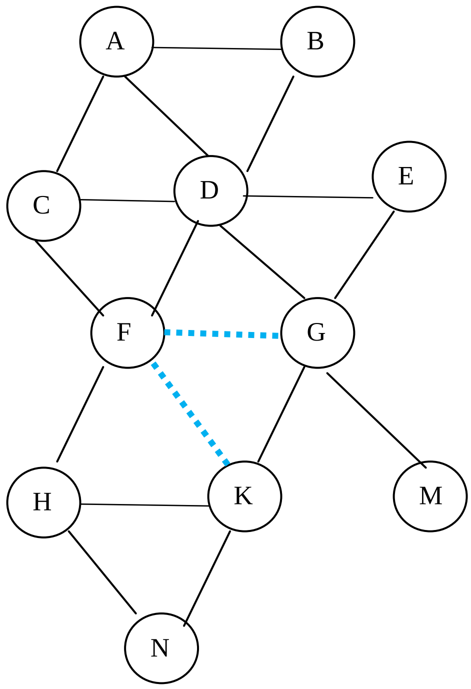
*Supprimer A de* $\mathcal{G}'_{\mathcal{M}}$

***Si*** $C \not\subset$ *cluster-set* ***Alors*** *cluster-set* $:=$ *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* := ∅ ;

$C_i$ := *Cluster-set[i]* ;

**Pour** *j* := 1 **à** (*m*-1) **Faire**

$C_j$ := *Cluster-set[j]* ;

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

*Insérer* $S_{ij}$ *dans Separator-set* ;

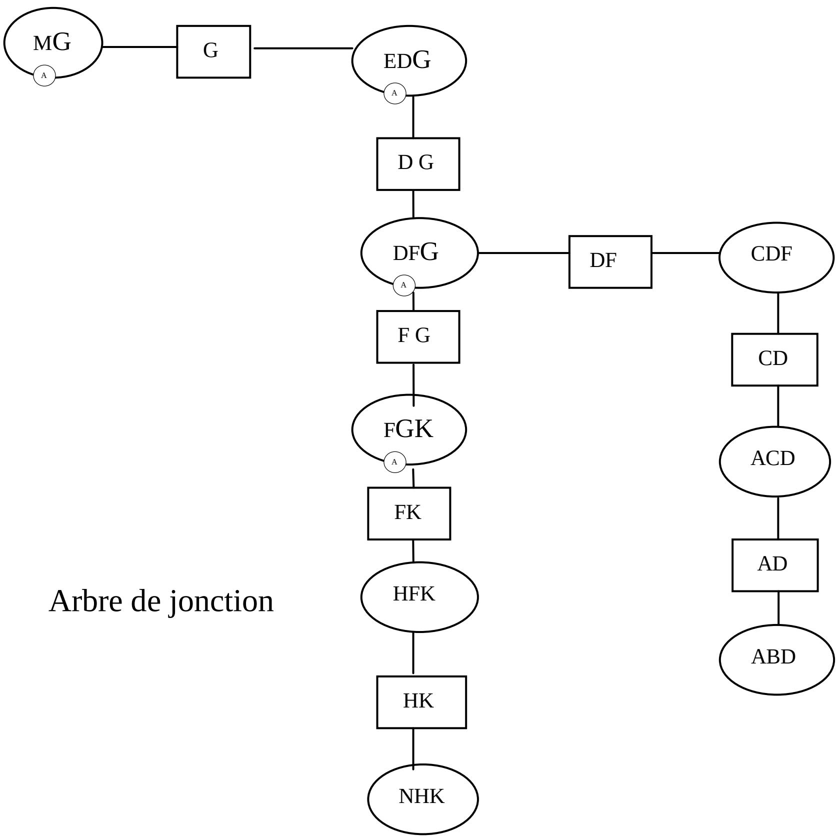
*Sélectionner un séparateur* $S_{ij}$ *à partir de Separator-set, en accord avec le critère de selection spécifié ci-après* ;

*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 | | 0 | 0 | |
| | EDG | G | 1 | 2 | * |
| | NHK | | 0 | 0 | |
| | ACD | | 0 | 0 | |
| | CDF | | 0 | 0 | |
| | DFG | G | 1 | 2 | |
| | HFK | | 0 | 0 | |
| | FGK | G | 1 | 2 | |
| ABD | EDG | D | 1 | 2 | |
| | NHK | | 0 | 0 | |
| | ACD | AD | 2 | 4 | * |
| | CDF | D | 1 | 2 | |
| | DFG | D | 1 | 2 | |
| | HFK | | 0 | 0 | |
| | FGK | | 0 | 0 | |
| EDG | NHK | | 0 | 0 | |
| | ACD | | 0 | 0 | |
| | CDF | D | 1 | 2 | |
| | DFG | DG | 2 | 4 | * |
| | HFK | | 0 | 0 | |
| | FGK | G | 1 | 2 | |
| NHK | ACD | | 0 | 0 | |
| | CDF | | 0 | 0 | |
| | DFG | | 0 | 0 | |
| | HFK | HK | 2 | 4 | * |
| | FGK | K | 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 | * |

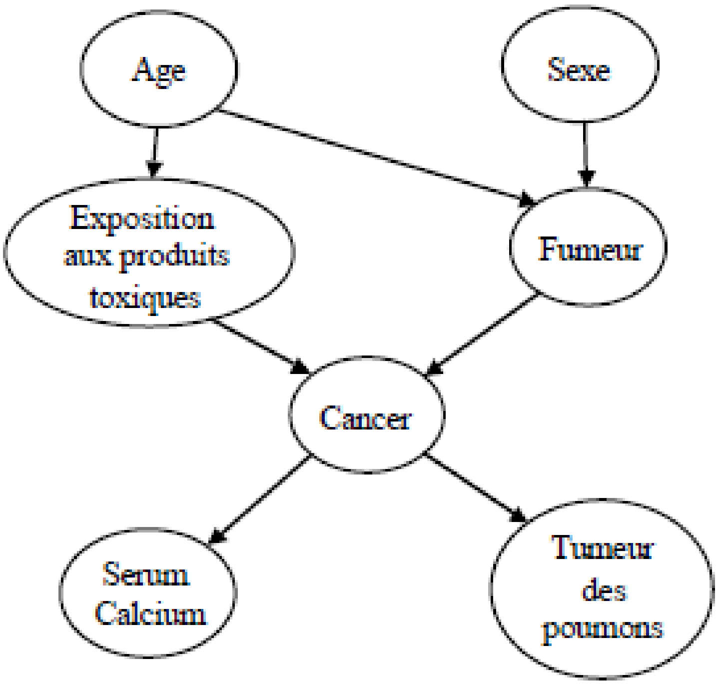



Arbre de jonction

**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.

- Cas d'un réseau Bayésien

– 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 :

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

- Cas des réseaux possibilistes :

- La composante graphique est identique
- Une composante numérique : qui consiste à quantifier les différents liens représentés par le graphe en utilisant les distributions possibilistes conditionnelles de chaque nœud dans le contexte de ses parents. Ces dernières doivent obéir à la condition de normalisation.

  Pour chaque variable A :
  - Si $U_A = \emptyset$ (A est un nœud racine), alors les possibilités a priori relatives à la variable A doivent satisfaire :

$$max_a \Pi(a) = 1, \forall a \in D_A,$$

  - Si $U_A \neq \emptyset$, alors les distributions conditionnelles de la variable A dans le contexte de ses parents doivent satisfaire :

$$max_a \Pi(a \mid a_A) = 1, \forall a \in D_A, u_A \in D_{U_A}$$

La différence de la définition du conditionnement a conduit à définir deux types de réseaux causaux possibilistes :

- Cas d'un réseau possibiliste basé sur le min

$$\pi(\omega \mid_m \phi) = \begin{cases} 1 & \text{si } \pi(\omega) = \Pi(\phi) \text{ et } \omega \models \phi \\ \pi(\omega) & \text{si } \pi(\omega) < \Pi(\phi) \text{ et } \omega \models \phi \\ 0 & \text{sinon} \end{cases}$$

- Cas d'un réseau possibiliste basé sur le produit

**(Conditionnement possibiliste basé sur le produit)**

$$\pi(\omega \mid_p \phi) = \begin{cases} \frac{\pi(\omega)}{\Pi(\phi)} & si\ \omega \models \phi \\ 0 & sinon \end{cases}$$

c- Comment se fait le calcul :

- des probabilités jointes.

**(Règle de chaînage)**

$$P(A_1, ..., A_N) = \prod_{i=1..N} P(A_i \mid U_{A_i})$$

$$\boxed{\begin{aligned} &\mathrm{P}(A,S,E,F,C,T,SC) = \\ &\quad \mathrm{P}(A)\,.\,\mathrm{P}(S)\,. \\ &\quad \mathrm{P}(E \mid A)\,.\,\mathrm{P}(F \mid A,S)\,. \\ &\quad \mathrm{P}(C \mid E,F)\,. \\ &\quad \mathrm{P}(SC \mid C)\,.\,\mathrm{P}(T \mid C) \end{aligned}}$$

Exemple :

P(a∧¬s∧e∧f∧¬c∧l∧t)=P(a)*P(¬s)*P(e|a)*P(f|¬s)*P(¬c| e∧f)*P(l|¬c)*P(t|¬c)

- Cas possibiliste basé sur le min:

**(Règle de chaînage basée sur le minimum)**

$$\pi_m(A_1, ..., A_N) = \min_{i=1..N} \Pi(A_i \mid U_{A_i})$$

Exemple :

Π(a∧¬s∧e∧f∧¬c∧l∧t)= min(Π(a), Π (¬s),Π (e|a), Π (f|¬s), Π (¬c| e∧f),Π (l|¬c), Π (t|¬c))

- Cas possibiliste basé sur le produit:

**(Règle de chaînage basée sur le produit)**

$$\pi_p(A_1, ..., A_N) = \prod_{i=1..N} \Pi(A_i \mid U_{A_i})$$

Exemple :

Π (a∧¬s∧e∧f∧¬c∧l∧t)= Π (a)* Π (¬s)* Π (e|a)* Π (f|¬s)* Π (¬c| e∧f)* Π (l|¬c)* Π (t|¬c)

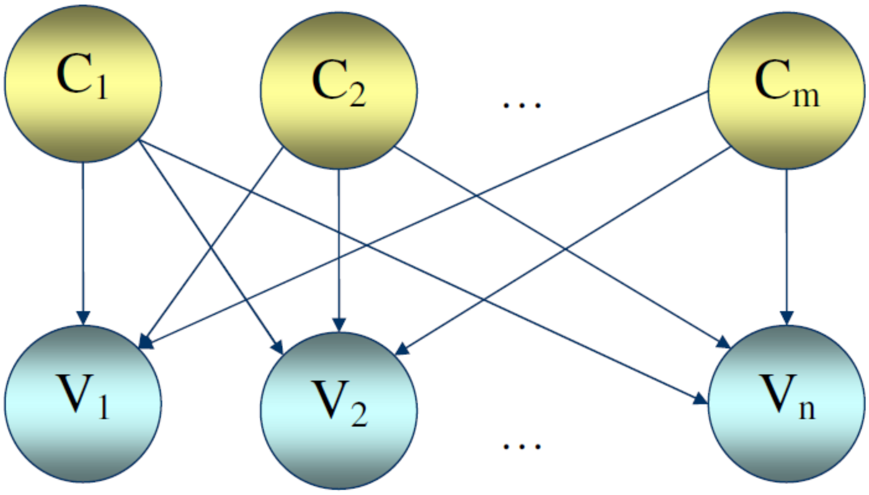
**Exercice 3:**

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

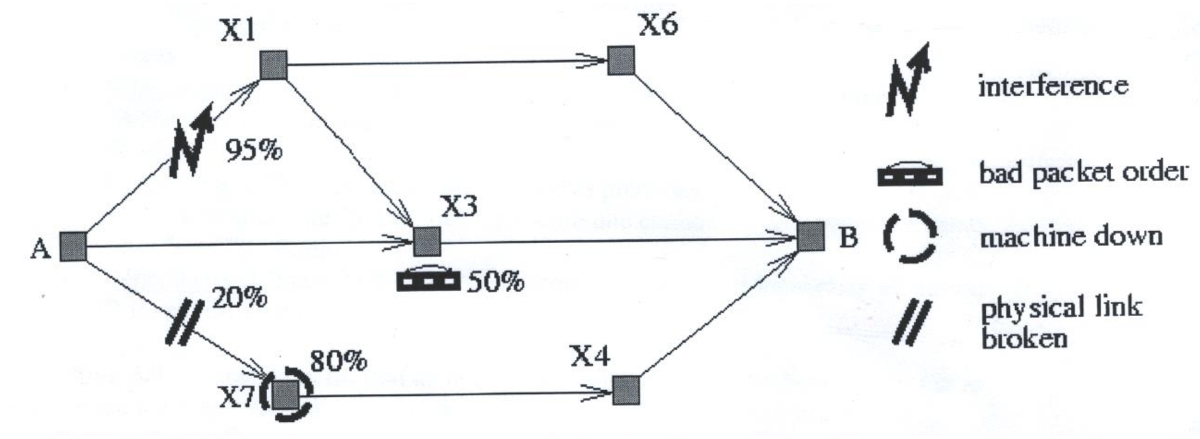
Proposez une structure du réseau Bayésien.

Chaque document est représenté par un vecteur de caractéristiques $X(x_{1x},., x_{mx})$. Les caractéristiques sont les termes qui peuvent bien identifier les classes de documents.

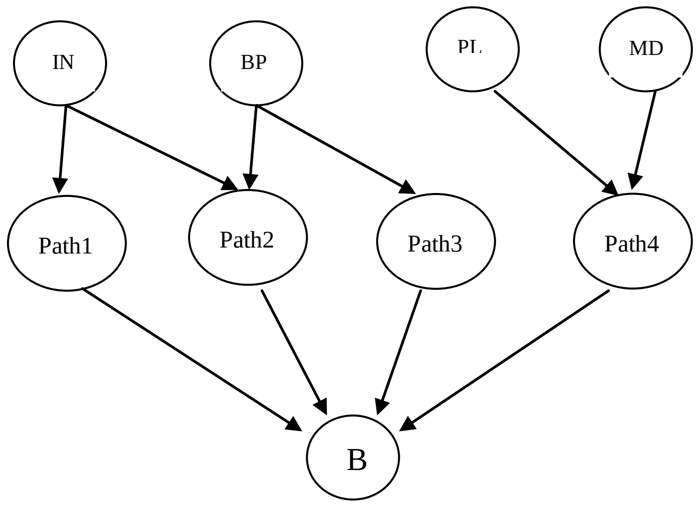
Le réseau bayésien utilisé dans ce système est un réseau très simple qui contient de 2 couches, il se base sur une hypothèse que les caractéristiques de documents sont indépendantes. Dans ce réseau les nœuds $V_i$ représentent les caractéristiques et les nœuds $C_j$ représentent les classes :



**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°4**

**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(variable d'intérêt | 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.