Étapes pour bâtir un réseau bayésien

- Comment bâtir un réseau bayésien afin de modéliser un environnement/problème donné ?

- On a besoin de spécifier 2 choses :
  - la structure du réseau
    (quelles indépendances peut-on supposer?)
  - les tables de probabilités
    (quelle est la relation entre les variables de l'environnement ?)

Spécifier la structure d'un RB

- L'approche la plus simple pour construire la structure du réseau est de le faire à la main

1. Choisir un ordre des variables $X_1, \ldots ,X_n$
2. Pour i = 1 to n :
   I. ajouter X_i au réseau
   II. choisir les parents $X_1, \ldots ,X_{i-1}$ tels que $P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \ldots X_{i-1})$
   III. ce choix garantit que :

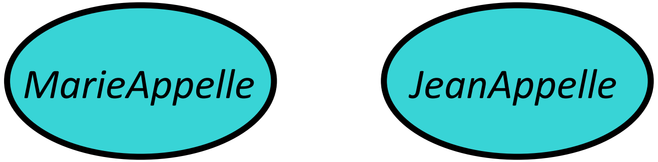
$$
\begin{aligned}
P(X_1, \ldots ,X_n) &= \Pi_{i=1}^{n} P(X_i \mid X_1, \ldots , X_{i-1}) \quad (\textit{règle de chainage}) \\
&= \Pi_{i=1}^{n} P(X_i \mid Parents(X_i)) \quad (\text{par construction})
\end{aligned}
$$

- Pour construire un bon RB, sa structure doit refléter les indépendances conditionnelles du problème

- Dans quel ordre ajouter les nœuds au réseau?
  - mettre les « causes racines » d'abord, ensuite les nœuds qu'ils influencent directement

Exemple

- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*

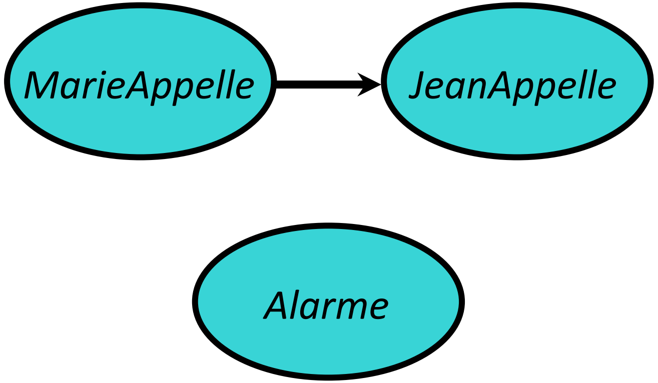


- $P(J|M) = P(J)$?

Hugo Larochelle et Froduald Kabanza

Exemple

- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*

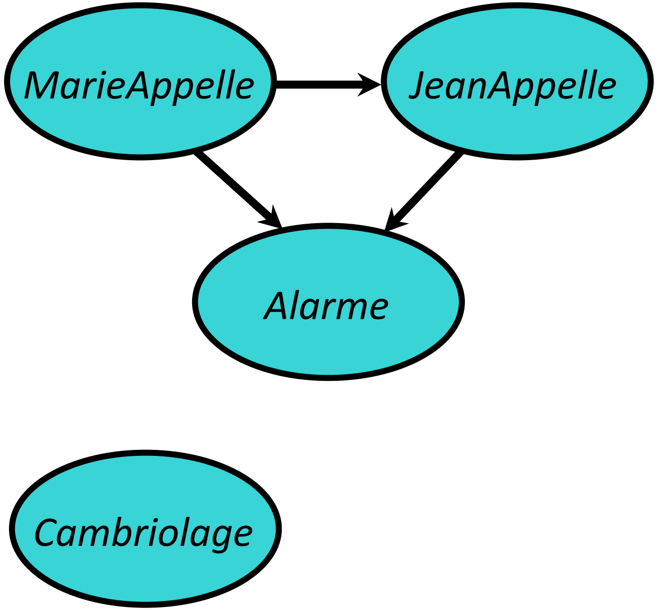


- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$?

Exemple

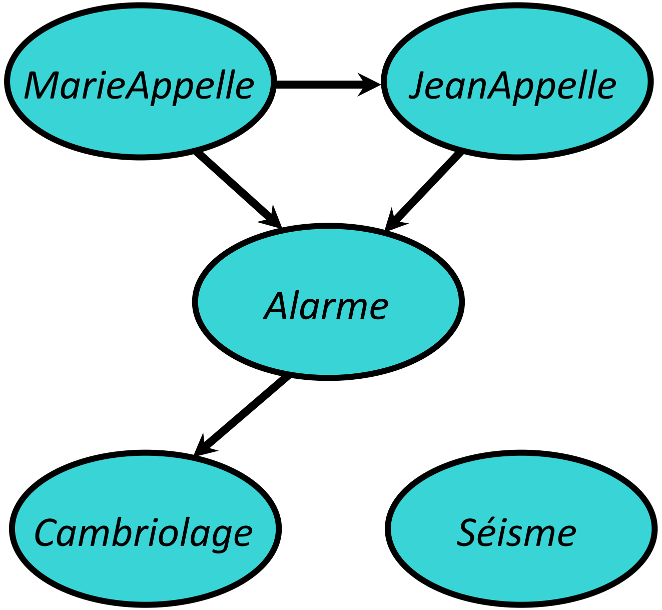
- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*

- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$?
  $P(C|A,J,M) = P(C)$?



Exemple

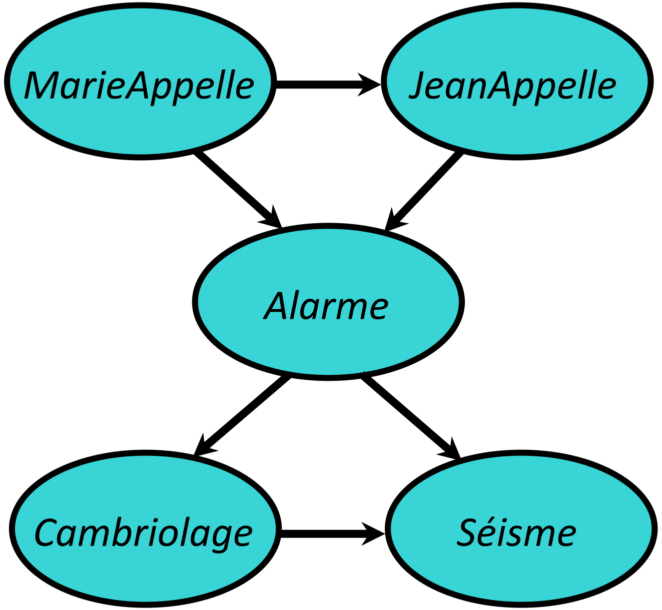
- Supposons qu'on ordonne les variables comme suit : *M, J, A, C, S*



- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
  $P(C|A,J,M) = P(C)$? **Non**
- $P(S|C,A,J,M) = P(S|A)$?
  $P(S|C,A,J,M) = P(S|A,C)$?

Exemple

- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*



- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
  $P(C|A,J,M) = P(C)$? **Non**
- $P(S|C,A,J,M) = P(S|A)$? **Non**
  $P(S|C,A,J,M) = P(S|A,C)$? **Oui**

Exemple

- Déterminer l'indépendance conditionnelle est très difficile dans le sens non causal
  - par exemple, en médecine, des études ont démontré que les experts préfèrent donner des probabilités dans le sens causal (pathologie → symptôme) plutôt que dans le sens diagnostique

- Un réseau avec des dépendances diagnostiques (effet → cause) est généralement moins compacte
  - dans le cas présent : 1 + 2 + 4 + 2 + 4 = 13 nombres pour représenter les tables de probabilité conditionnelle du réseau au lieu de 10 pour la première version

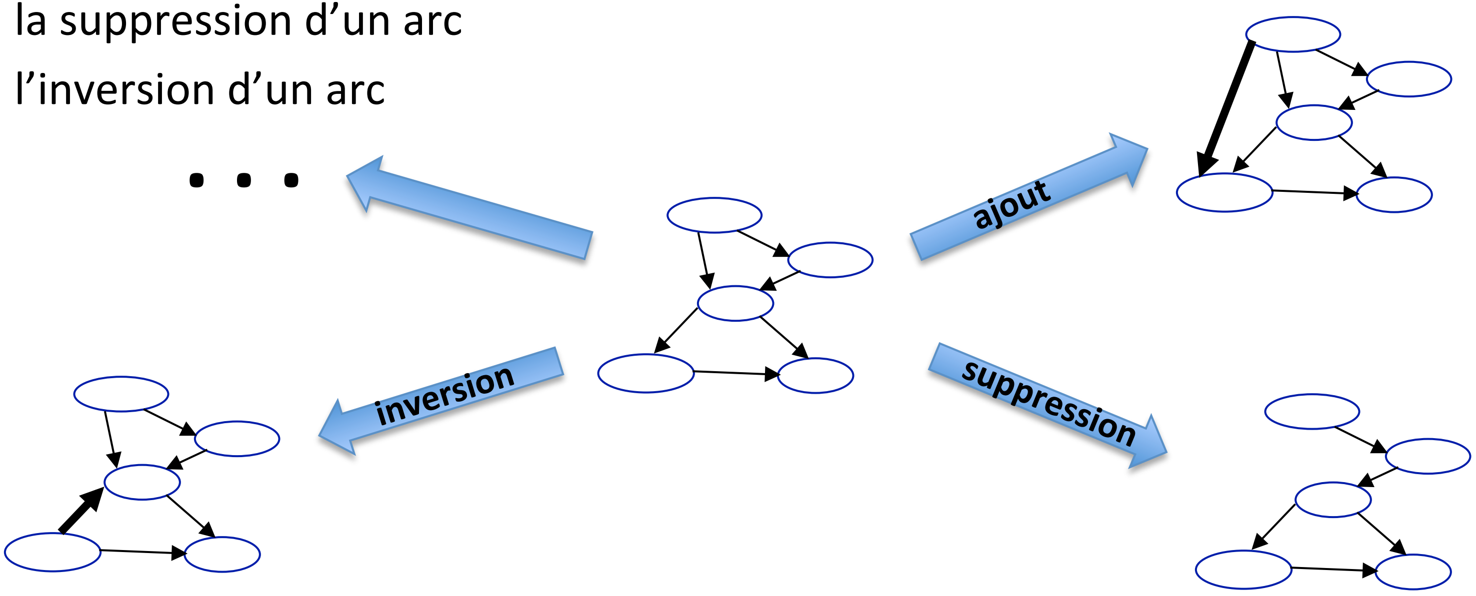
Spécifier la structure d'un RB

- Quoi faire si on n'a pas accès à un expert pour définir un bon graphe de RB ?

- On peut aussi tenter d'obtenir la structure du RB à partir de données, à l'aide de la recherche locale (par exemple *Hill Climbing*) :
  1. on débute avec un **graphe acyclique aléatoire** comme graphe courant
  2. on **obtient ses tables de probabilités** à partir des fréquences d'observation du graphe courant
  3. on utilise la recherche locale pour **générer des graphes successeurs** du graphe courant
     I. on obtient les tables de probabilités du graphe successeur
     II. on remplace le graphe courant par le successeur s'il est « meilleur »
  4. on retourne à 2. jusqu'à un certain critère d'arrêt

Spécifier la structure d’un RB

- On génère des successeurs à partir des modifications de graphe suivantes
  - l’ajout d’un arc
  - la suppression d’un arc
  - l’inversion d’un arc



Spécifier la structure d'un RB

- La fonction objectif à maximiser par la recherche locale est :

$$\underbrace{\Sigma_t \log\ P(X_1 = x_1^t,\ldots,X_n = x_n^t)}_{\textbf{log probabilité des données}} \ - \ \underbrace{M\ (\log T)\ /\ 2}_{\textbf{complexité du graphe}}$$

  - $\{x_1^t,\ldots, x_n^t\}$ est la $t^{\text{ième}}$ donnée de mon ensemble de T données
  - M est le nombre de paramètres requis par les tables de probabilités conditionnelles du réseau bayésien

- On cherche donc un graphe
  - qui **explique bien les données** (leur donne une haute probabilité)
  - qui est **compacte** (qui a peu de paramètres)
- Pour en savoir plus : voir section 20.2.5