

Cours Graphes et Applications

Partie II – « Graphes Aléatoires »

Alexandre Aussem

aaussem@univ-lyon1.fr

Thème de recherche : Conception de Modèles pour l'Aide à la Décision



PRISMa
Université Lyon 1



Plan du cours

- Les Graphes et l'Optimisation Combinatoire
- Les Graphes comme modèles de représentation de connaissances incertaines.
- Les **Modèles Graphiques** pour représenter des distributions de probabilités
- L'inférence dans les **Réseaux Bayésiens** pour le **diagnostic** des systèmes complexes.
- L'apprentissage des Réseaux Bayésiens
- **Applications**

La Théorie des Graphes

- Formalisme puissant pour modéliser les systèmes :
 - **Noeuds** = entités du système, e.g. machines, routeurs, services, pages Web, pixels,
 - **Arcs** = contraintes, liens, communications, interactions
- Support à la résolution des nombreux problèmes d'**Optimisation Combinatoire**
 - Plus courts chemins, coloration, flots de valeur maximum, couplage maximum, coupes de capacité minimum, affectation etc.
- Applications “classiques”
 - Transports, télécom, réseaux, image, études de pannes,
- Applications non usuelles : modélisation dans un univers incertain
 - **Noeuds** = variables aléatoires ou états possibles
 - **Arcs** = dépendances probabilistes ou transitions aléatoires

La Théorie des Graphes

- La majorité des problèmes de reconnaissance en Optimisation Combinatoire sont dits NP-Complets
- NP-Complet : classe d'équivalence des problèmes NP selon la relation transitive « transformation polynomiale »
- NP : toute solution proposée se vérifie en temps polynomial

Voyageur de Commerce, $G=\{E,V\}$

- Formulation sous la forme d'un Programme Linéaire en Nombres Entiers (PLNE) NP-Complet :

$$\left\{ \begin{array}{l} \text{Min} \quad \sum_{ij} x_{ij} d_{ij} \\ \sum_i x_{ij} = 1, \forall j \in V \\ \sum_j x_{ij} = 1, \forall i \in V \\ \sum_{(i,j) \in X} x_{ij} < \text{card}(X), \forall X \subsetneq V \\ x_{ij} \in \{0,1\} \end{array} \right.$$

Coloration du Graphe $G=\{E,V\}$

- Formulation sous la forme d'un Programme Linéaire en Nombres Entiers (PLNE) NP-Complet :

$$\left\{ \begin{array}{l} \text{Min} \quad z \\ x_i \leq z \\ x_i \leq x_j - 1 + ny_{ij}, \forall (i, j) \in E \\ x_j \leq x_i - 1 + n(1 - y_{ij}), \forall (i, j) \in E \\ x_i \in \mathbb{N}, \forall i \in V \\ y_{ij} \in \{0,1\}, \forall (i, j) \in E \end{array} \right.$$

Résolution

- Outre les parcours de graphes, d'autres techniques existent pour résoudre les Pb d'Optimisation Combinatoire :
 - Méthodes (exactes) par séparation et évaluation « **Branch and Bound** » (technique de décomposition arborescente avec élagage)
 - Méthodes (exactes) de coupes « **Branch and Cut** » basées sur le simplexe et la génération itératives de contraintes
 - La **programmation dynamique** (exacte) utilise un principe d'optimalité exprimée sous une forme récursive
 - Méthodes **heuristiques** (approximatives en temps raisonnable)
 - Algorithmes gloutons (*greedy algorithms*)
 - Recuit simulé (*simulated annealing*)
 - Méthode tabou
 - Algorithmes génétiques, simulation d'essaims, fourmis etc.

Les Modélisation de l'Incertain

- Représenter exhaustivement des distributions multi-dimensionnelles est illusoire
 - Le nombre de paramètres croît exponentiellement avec le nombre de variables aléatoires
- Solution (début 90)
 - Modèles de distributions représentés par des graphes : les **Modèles Graphiques**

Les Modèles Graphiques

- Ce sont des modèles probabilistes novateurs pour la représentation des connaissances, fondés sur une description graphique des variables aléatoires.
- Idée : prendre en compte les *dépendances* et *indépendances conditionnelles* entre les variables
- Objectif : représenter des distributions multi-dimensionnelles de grande taille en évitant l'explosion combinatoire (complexité temporelle et spatiale)

Les Modèles Graphiques

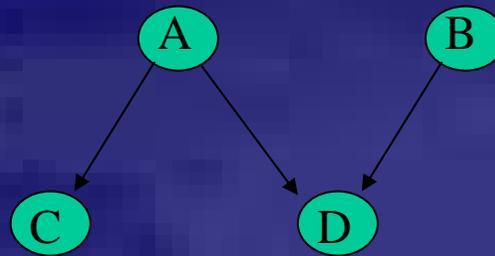
- Deux grandes classes :
 - Les **Réseaux Bayésiens**
 - Représentation asymétrique des dépendances
 - Modélise bien les relations causales (diagnostic)
 - Les **Champs de Markov**
 - Représentation symétrique des dépendances
 - Souvent utilisées pour modéliser les dépendances spatiales (analyse d'image)

Les Réseaux Bayésiens

- Représentent et encodent de façon compacte les distributions conjointes de variables aléatoires
- Exploitent les dépendances et indépendances conditionnelles pour éliminer les paramètres superflus
- Si **A** et **B** sont indépendants
$$P(\mathbf{A}, \mathbf{B}) = P(\mathbf{A})P(\mathbf{B})$$
- Si **A** et **B** sont indépendants *conditionnellement* à **C**
$$P(\mathbf{A}, \mathbf{B} | \mathbf{C}) = P(\mathbf{A} | \mathbf{C})P(\mathbf{B} | \mathbf{C})$$

Les Réseaux Bayésiens

- Modèles de représentation des connaissances, fondés sur une description graphique des variables aléatoires : **Directed Acyclic Graph (DAG)**



- Les nœuds sont les variables aléatoires et les arcs sont les relations (si possibles) causales entre ces variables
- L'absence d'arc signifie une indépendance conditionnelle

Tables de probabilités

- Dans chaque nœud, on stocke la table de probabilités conditionnelles locale $P(X_i|Pa_i)$ pour chaque configuration des parents Pa_i du nœud X_i

$$P(D|A,B)$$

| A | B | True | False |
|-------|-------|------|-------|
| False | False | 0.4 | 0.6 |
| False | True | 0.1 | 0.9 |
| True | False | 0.7 | 0.3 |
| True | True | 0.6 | 0.4 |

Les Réseaux Bayésiens

- On dira que le couple $\{G,P\}$ est un Réseau Bayésien, avec $G=\{V,E\}$ un DAG, s'il vérifie la **condition de Markov** : chaque variable X dans V est indépendante de ses non descendantes (ND_X) dans G conditionnellement à ses parents :

$$Ind_P(X, ND_X / Pa_i)$$

où Pa_i désigne l'ensemble des parents de X_i dans G .

- La condition de Markov implique la factorisation de la loi jointe :

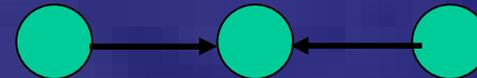
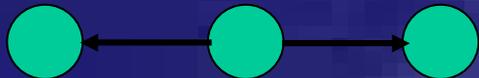
$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i / Pa_i)$$

Les Réseaux Bayésiens

- Cette propriété importante montre qu'il suffit de stocker les valeurs de $P(X_i|Pa_i)$ pour toutes les valeurs de X_i et les possibles instanciations conjointes de Pa_i dans une table de probabilités.
- Toute requête portant sur une ou plusieurs variables d'intérêt conditionnellement à d'autres (les observations partielles) peuvent être obtenue par inférence.

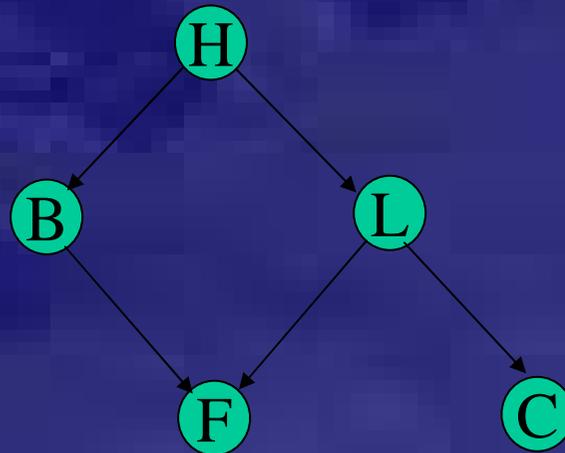
Les Réseaux Bayésiens

- Plusieurs types de chaînes sont possibles entre 3 variables :



Connexion convergente

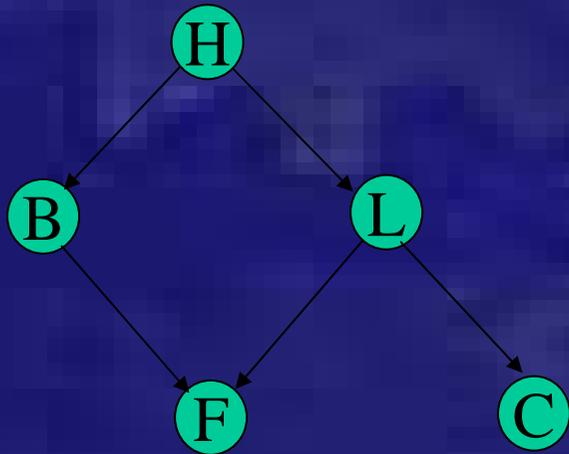
Exemple Illustratif



$$P(F,C,B,L,H) = P(F|B,L)P(C|L)P(B|H)P(L|H)P(H)$$

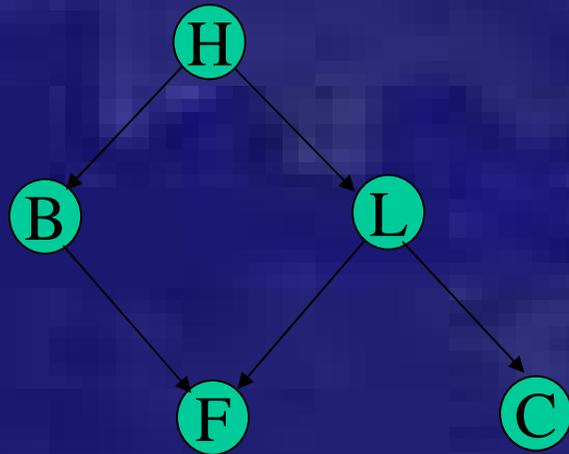
La structure du graphe implique la décomposition !

Exemple Illustratif



| Noeud | Parents | Indépendance Conditionnelle |
|-------|---------|-------------------------------|
| C | {L} | $I_p(C, \{H, B, F\} \{L\})$ |
| B | {H} | $I_p(B, \{L, C\} \{H\})$ |
| F | {B, L} | $I_p(F, \{H, C\} \{B, L\})$ |
| L | {H} | $I_p(L, \{B\} \{H\})$ |

Complexité spatiale



Taille mémoire de la loi conjointe :
 $2^5 - 1 = 31$

Taille mémoire du réseau bayésien :
 $1 + 2 + 2 + 4 + 2 = 11$

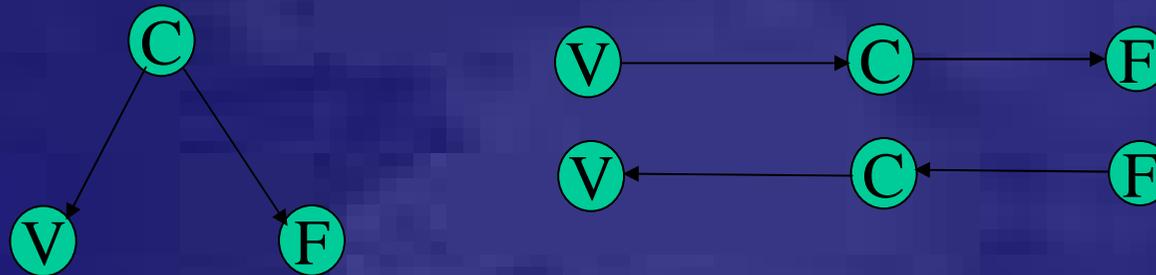
Dans le premiers cas, taille exponentielle, dans le second, taille linéaire avec le nombre de noeuds !

Illustration

1 1 2 2 2 2 1 2 2 1 2 1 2

- P uniforme. $\text{Ind}_P(V, F|C)$ mais pas $\text{Ind}_P(C, F)$

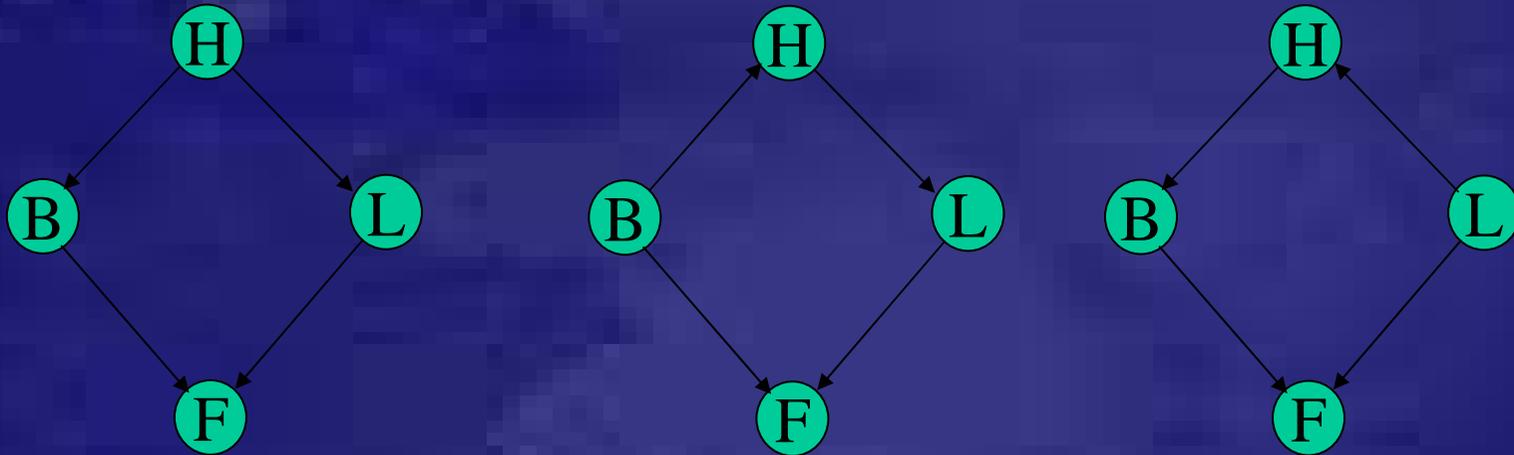
$$P(V, C, F) = P(C)P(V|F)P(F|C) = P(V)P(C|V)P(F|C) = P(F)P(C|F)P(V|F)$$



P vérifie la condition de Markov avec ces 3 DAG; ils encodent la même distribution de probabilité !

Equivalence de Markov

- Les DAG qui encodent les mêmes d-séparations forment une classe d'équivalence



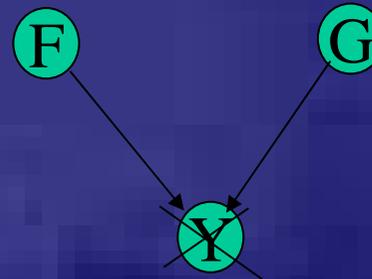
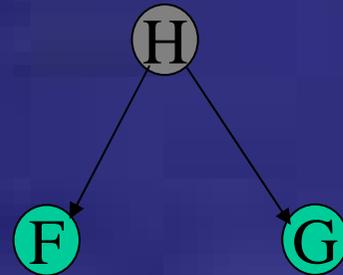
Impossible de les distinguer à partir des données !

La causalité

- On cherche idéalement à trouver les causes des phénomènes observés
- La causalité est une notion sujette à caution, elle fait l'objet de controverses depuis des siècles, Hume 1748, Piaget 1966 etc..
- Définition “**opérationnelle**” : Si une manipulation particulière de X , en forçant sa valeur, provoque la modification de la distribution de Y , alors X cause Y .
- On suppose ici que cause et effets sont corrélés

Corrélation et causalité

Si F et G sont corrélés, F est-il la cause de G ?



La causalité

- F = Finasteride (médicament) et G = Hair Growth. Une population d'individus est observée.
- Cas 2 - Un autre produit X a eu de l'effet et suscite l'intérêt de l'individu pour F
- Cas 3 – F a de l'effet et suscite un intérêt pour F en retour.
- Cas 4 – H est l'inquiétude de l'individu, lequel prend 2 produits F et X. Seul X a un effet sur G.
- Cas 5 – F et G suscitent une hypertension Y. Or, la population observée a de l'hypertension. Y est alors instanciée (biais de sélection).

Contraintes sur le graphe

- La d-séparation est un critère important qui permet de caractériser graphiquement toutes les contraintes d'indépendance des lois P qui peuvent être représentées par une même DAG.
- Il faut introduire la notion de graphes :
 - Chaînes & chemins, simples ou composés,
 - Parents, enfants, descendants, ancêtres etc.
- Par une chaîne de X vers Y transite une information bruitée : les sommets sont des vannes ouverts ou fermées.

D-séparation

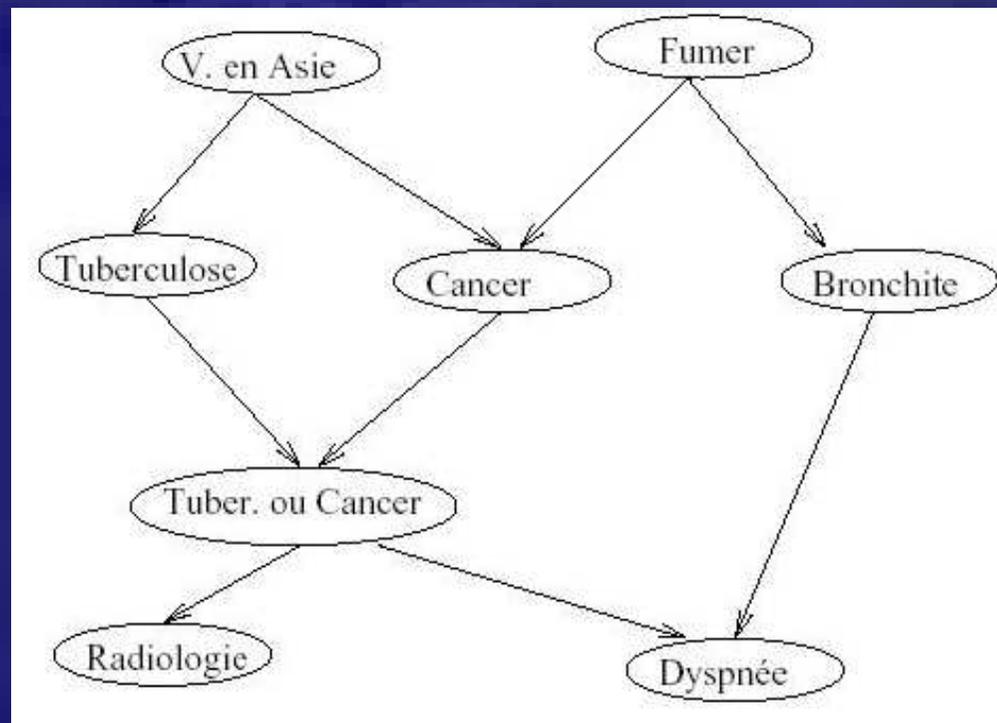
- Une chaîne est dite **ouverte** si toutes les vannes sont ouvertes auquel cas la chaîne laisse passer l'information.
- A l'inverse, si l'une des vanne est bloquée, la chaîne est dite **bloquée**.
- l'information qu'apporte X sur Y peut se voir comme la somme des **flots d'information** sur tous les chaînes ouvertes reliant X à Y.
- Mécanisme d'ouverture et de fermeture des vannes ?

D-séparation

- Formellement, un chaîne entre X et Y est **bloquée** par un ensemble de noeuds Z s'il existe un noeud W sur cette chaîne vérifiant l'une des conditions :
 - W n'est pas un noeud convergent : W est dans Z
 - W est un noeud convergent : ni W , ni aucun de ses descendants ne sont dans Z .
- Deux noeuds X et Y sont dits *d-séparés* par Z dans le graphe G , $Dsep_G(X;Y|Z)$, si tous les chaînes (simples) entre X et Y sont bloquées par Z .
- La d-séparation dresse un parallèle élégant entre l'algorithmique des graphes et le calcul des indépendances conditionnelles dans une distribution de probabilités.

Illustration

Réseau de la dyspnée : *Asia*



Inférence & Diagnostic

Information  Connaissance

- Permet d'**inférer des connaissances**, e.g. $P(\mathbf{D}|\mathbf{C},\mathbf{B})$, à partir d'observations partielles, bruitées etc.
- Représentation graphique des connaissances lisible par les non spécialistes (e.g., médecins)
- Autorise des requêtes probabilistes du type : Quelle probabilité de telle maladie sachant tels symptômes.
- Permet de hiérarchiser les diagnostics (simples ou multiples) en fonction des probabilités *a posteriori* .

Schémas d'Inférence

- $P(\mathbf{D}|\mathbf{B})$: calcul d'inférence causal
- $P(\mathbf{A}|\mathbf{D})$: calcul de diagnostic
- $P(\mathbf{A})$: calcul d'explication
- Problèmes NP-difficiles pour des ensembles de variables quelconques

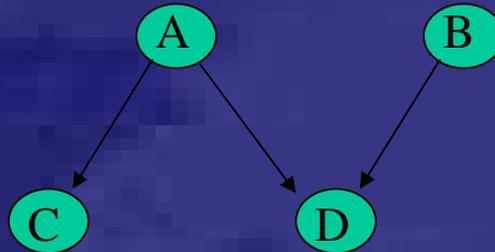
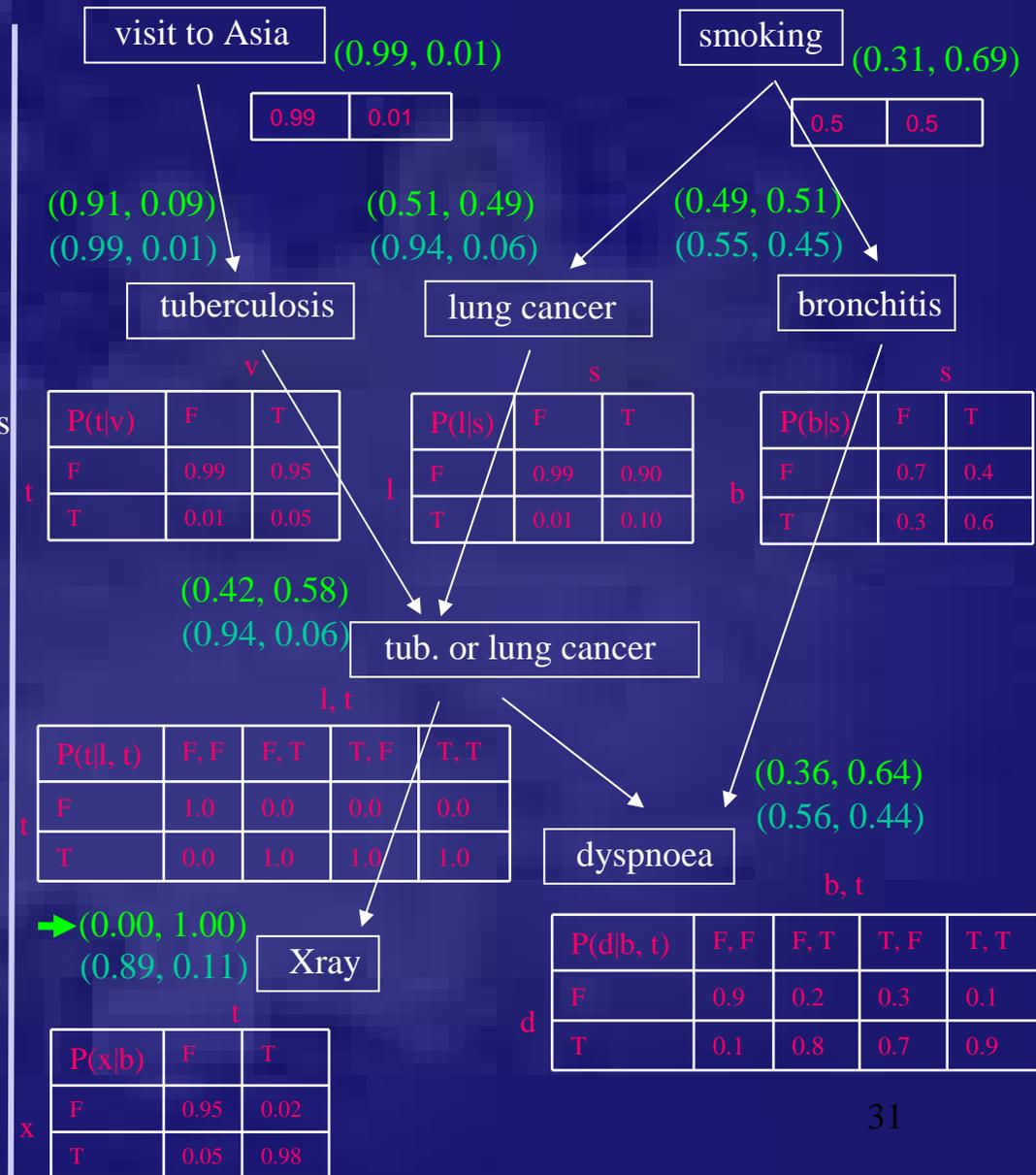


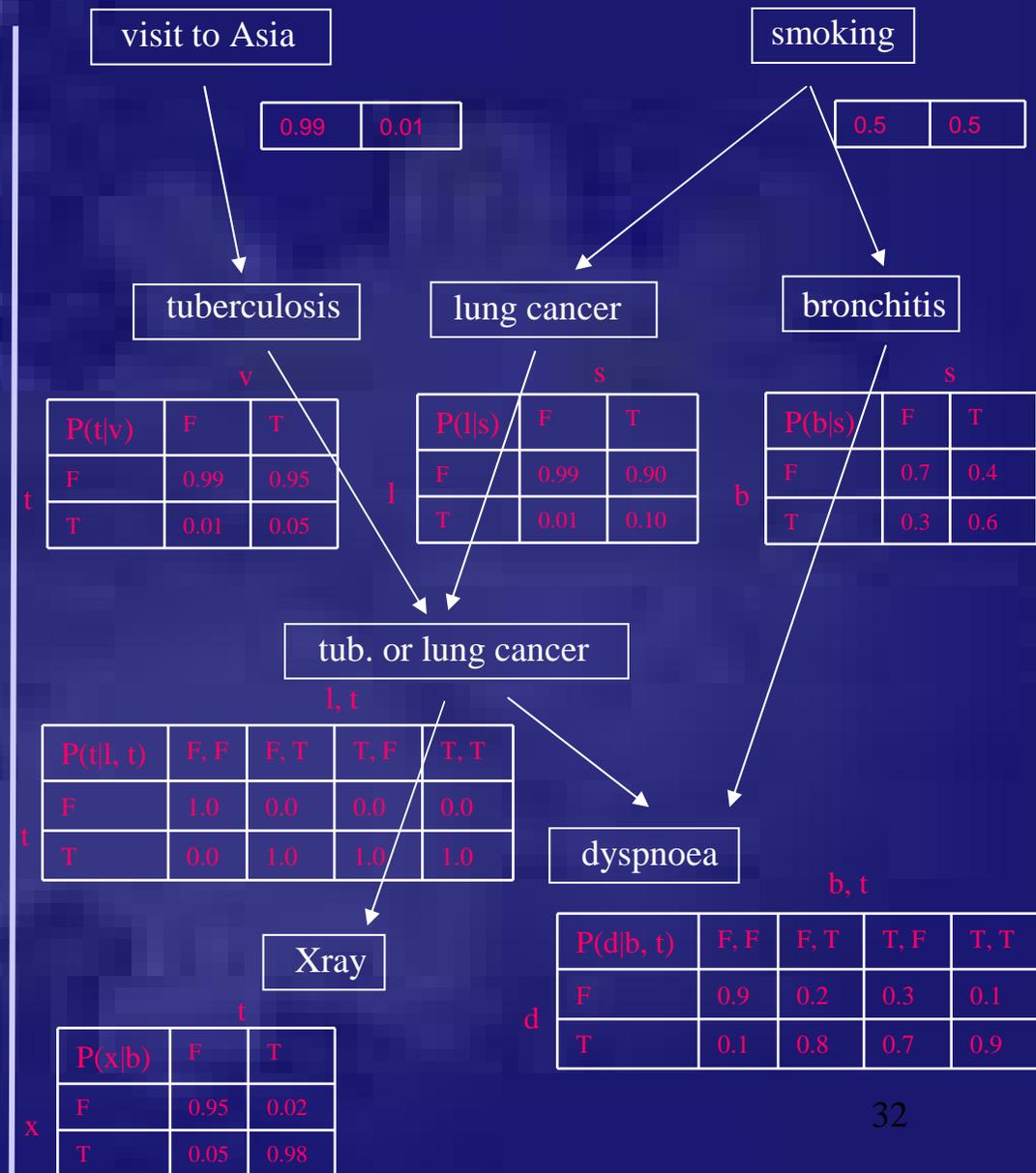
Illustration : Diagnostic Médical

- Support formel
 - = Graphe orienté acyclique
 - + Tables de probabilités conditionnelles
 - = Réseau Bayésien
- État de connaissance courant
 - = Distributions de probabilités, une par variable, dérivées des tables de probabilités conditionnelles
- Prise en compte d'observations en vue d'établir un diagnostic
 - => Injection de nouvelles connaissances, ex., la radiographie des poumons est positive
 - => Perturbation de l'état de connaissance
 - => Retour à l'équilibre de l'état de connaissance (inférence)
 - => **Nouvel état de connaissance**
 - => Utilisation de l'état courant ou du différentiel par rapport aux états précédents pour établir un diagnostic



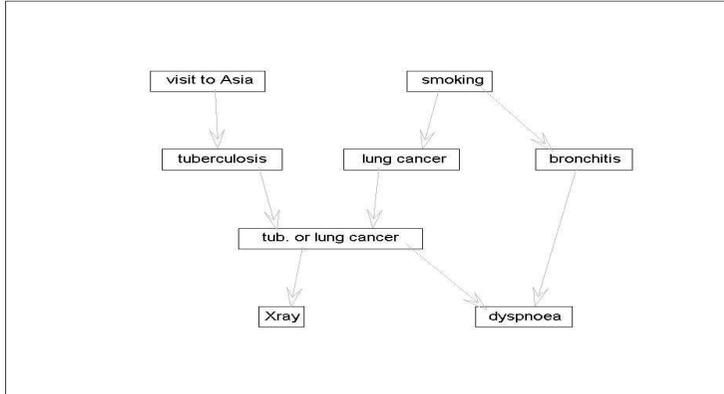
Comment faire l'apprentissage du modèle ?

- Construction du modèle graphique
 - Identification des variables et des domaines de valeurs qui leur sont associés
 - Construction du graphe causal à l'aide d'un expert du domaine ou par **apprentissage automatique des causalités**
- Construction par **apprentissage** des tables de probabilités conditionnelles
- Validation du modèle : par des experts et/ou par des techniques statistiques.
- Construction de l'interface graphique conviviale
 - Définition des fonctionnalités
 - Développement logiciel



Comment faire un diagnostic?

Simulation (prototype construit et résultats obtenus avec le logiciel numérique : Matlab + toolbox BNT)



evidence setting:
U, T, U, U, U, U, T, U

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.98782 | 0.012185 |
| smoking : | 0 | 1 |
| tuberculosis : | 0.93282 | 0.067183 |
| lung cancer : | 0.35401 | 0.64599 |
| bronchitis : | 0.4 | 0.6 |
| tub. or lung cancer : | 0.29354 | 0.70646 |
| Xray : | 0 | 1 |
| dyspnoea : | 0.26806 | 0.73194 |

evidence setting:
T, U, U, U, U, U, T, U

| | NO | YES |
|-----------------------|---------|---------|
| visit to Asia : | 0 | 1 |
| smoking : | 0.36299 | 0.63701 |
| tuberculosis : | 0.66228 | 0.33772 |
| lung cancer : | 0.62851 | 0.37149 |
| bronchitis : | 0.5089 | 0.4911 |
| tub. or lung cancer : | 0.30937 | 0.69063 |
| Xray : | 0 | 1 |
| dyspnoea : | 0.3189 | 0.6811 |

evidence setting:
U, U, U, U, U, U, U, U

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.99 | 0.01 |
| smoking : | 0.5 | 0.5 |
| tuberculosis : | 0.9896 | 0.0104 |
| lung cancer : | 0.945 | 0.055 |
| bronchitis : | 0.55 | 0.45 |
| tub. or lung cancer : | 0.93517 | 0.064828 |
| Xray : | 0.88971 | 0.11029 |
| dyspnoea : | 0.56403 | 0.43597 |

evidence setting:
U, U, U, U, U, U, U, T

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.98968 | 0.010325 |
| smoking : | 0.366 | 0.634 |
| tuberculosis : | 0.98115 | 0.018845 |
| lung cancer : | 0.89724 | 0.10276 |
| bronchitis : | 0.16603 | 0.83397 |
| tub. or lung cancer : | 0.87946 | 0.12054 |
| Xray : | 0.8379 | 0.1621 |
| dyspnoea : | 0 | 1 |

evidence setting:
U, T, U, U, U, U, U, T

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.98981 | 0.010193 |
| smoking : | 0 | 1 |
| tuberculosis : | 0.98457 | 0.015427 |
| lung cancer : | 0.85167 | 0.14833 |
| bronchitis : | 0.11984 | 0.88016 |
| tub. or lung cancer : | 0.83778 | 0.16222 |
| Xray : | 0.79914 | 0.20086 |
| dyspnoea : | 0 | 1 |

evidence setting:
T, U, U, U, U, U, U, T

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0 | 1 |
| smoking : | 0.37408 | 0.62592 |
| tuberculosis : | 0.91225 | 0.087751 |
| lung cancer : | 0.90047 | 0.099525 |
| bronchitis : | 0.1886 | 0.8114 |
| tub. or lung cancer : | 0.8177 | 0.1823 |
| Xray : | 0.78046 | 0.21954 |
| dyspnoea : | 0 | 1 |

evidence setting:
U, U, U, U, U, U, T, U

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.98684 | 0.013156 |
| smoking : | 0.31225 | 0.68775 |
| tuberculosis : | 0.90759 | 0.092411 |
| lung cancer : | 0.51129 | 0.48871 |
| bronchitis : | 0.49367 | 0.50633 |
| tub. or lung cancer : | 0.42396 | 0.57604 |
| Xray : | 0 | 1 |
| dyspnoea : | 0.35923 | 0.64077 |

evidence setting:
U, U, U, U, U, U, T, T

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.98602 | 0.013984 |
| smoking : | 0.21439 | 0.78561 |
| tuberculosis : | 0.88607 | 0.11393 |
| lung cancer : | 0.37875 | 0.62125 |
| bronchitis : | 0.31813 | 0.68187 |
| tub. or lung cancer : | 0.27127 | 0.72873 |
| Xray : | 0 | 1 |
| dyspnoea : | 0 | 1 |

evidence setting:
U, T, U, U, U, U, T, T

| | NO | YES |
|-----------------------|---------|----------|
| visit to Asia : | 0.9875 | 0.012496 |
| smoking : | 0 | 1 |
| tuberculosis : | 0.92473 | 0.075266 |
| lung cancer : | 0.27629 | 0.72371 |
| bronchitis : | 0.28629 | 0.71371 |
| tub. or lung cancer : | 0.20855 | 0.79145 |
| Xray : | 0 | 1 |
| dyspnoea : | 0 | 1 |

evidence setting:
T, U, U, U, U, U, T, T

| | NO | YES |
|-----------------------|---------|---------|
| visit to Asia : | 0 | 1 |
| smoking : | 0.29797 | 0.70203 |
| tuberculosis : | 0.60829 | 0.39171 |
| lung cancer : | 0.55573 | 0.44427 |
| bronchitis : | 0.37118 | 0.62882 |
| tub. or lung cancer : | 0.18623 | 0.81377 |
| Xray : | 0 | 1 |
| dyspnoea : | 0 | 1 |

Apprentissage des RB

- L'apprentissage de la **structure** à partir de données est NP-difficile : la taille de l'espace des DAG est super-exponentielle en fonction du nombre de variables.

Par exemple, $T(5) = 29281$

- L'apprentissage des **tables de probabilités** est aisé, simple calcul fréquentiel (polynomial)
- L'**inférence** exacte est NP-difficile, de même que l'inférence approchée...

Apprentissage du DAG

- Deux grandes familles de méthodes existent pour l'apprentissage du DAG :
 - celles fondées sur **la satisfaction de contraintes d'indépendance conditionnelle** entre variables
 - celles dites "bayésiennes" fondées sur la **maximisation d'un score** (BIC, MDL, BDe, etc.).
- Les méthodes sous contraintes sont **déterministes**, relativement rapides et bénéficient des critères d'arrêts clairement définis. Utilisent des informations statistiques dans les données (niveau de signification arbitraire). Les erreurs commises au début se répercutent en cascade.
- Les méthodes à base de score incorporent des probabilités a priori sur la structure du graphe, traitent plus facilement les **données manquantes**, mais sont facilement piégées dans les **minima locaux**. Le graphe final obtenu dépend des conditions initiales.

Inférence

- Technique d'envoi asynchrone de messages jusqu'à équilibre si le DAG est un arbre.
- Cut-set conditioning : instancier des variables pour que le graphe restant soit un arbre.
- Algorithme de l'arbre de jonction.
- Méthodes d'approximation : effectuent des échantillonnage de Gibbs sur des sous-ensembles de variables

Arbre de jonction

- Algorithme de l'arbre de jonction [Jensen90]
 - Phase de construction
 - Transforme le graphe en un arbre de jonction (arbre couvrant minimal) dont les nouveaux nœuds sont des clusters des nœuds initiaux.
 - 1 : Moralisation (marier tous les parents)
 - 2 : Triangulation et extraction de cliques (les voisins sont connectés deux à deux, ils appartiennent à des cliques qui formeront les nœuds du futur arbre de jonction)
 - 3 : Création d'un arbre couvrant minimal (arbre de jonction)
 - Propagation et mise à jour de la distribution

Applications des RB

- Biologie & Santé & Epidémiologie : p.ex. réseau de régulation entre gènes à partir des puces à ADN,
- Finance : analyse de risques de prêts, détection des mauvais payeurs.
- Vision : reconnaissance de visage, trafic routier
- Diagnostic de pannes/bugs/maladies : Microsoft, Intel, Ricoh...
- Traitement du langage
- Prévision, classification, datamining, marketing etc.