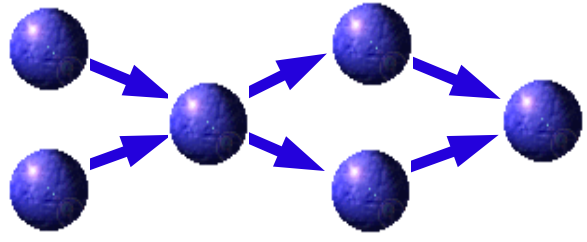
Réseaux Bayésiens

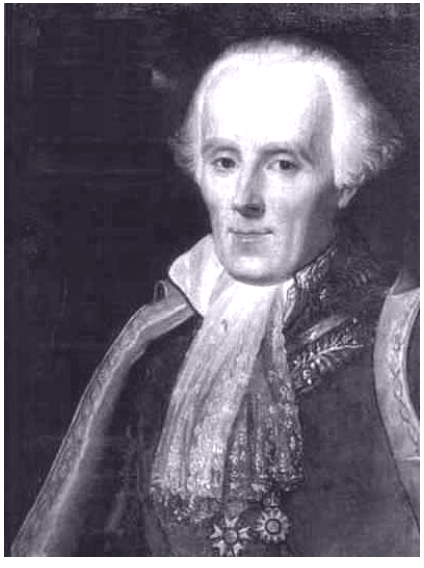
Formation CNRS-RISC – 3-4 novembre 2005

Philippe Leray
Philippe.Leray@insa-rouen.fr
http://asi.insa-rouen.fr/~pleray/

INSA Rouen – Laboratoire PSI (Perception, Systèmes, Information)



Un peu d'histoire...



- Rev. Thomas Bayes (1702–1761)
  - 1763 : *An Essay towards solving a Problem in the Doctrine of Chances*
    - probabilité conditionnelle :

      If there be two subsequent events, the probability of the second b/N and the probability of both together P/N, and it being first discovered that the second event has also happened, the probability I am right [i.e., the conditional probability of the first event being true given that the second has happened] is P/b.

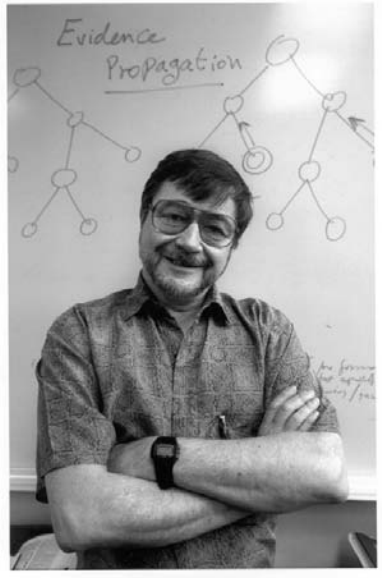
    - théorème de Bayes
- Pierre-Simon Laplace (1749–1827))
  - 1774 : *Mémoire sur la Probabilité des Causes par les Evénements*

Deux siècles plus tard...

RULE037

IF the organism
1) stains grampos
2) has coccus shape
3) grows in chains

THEN
There is suggestive evidence (.7) that the identity of the organism is streptococcus.



- 1970-1990 : L'ère des systèmes experts
  - base de règles
    `si X=vrai et Y=absent alors Z=faux`
  - moteur d'inférence (chainage avant, arrière)
- Judea Pearl (19xx–) : les réseaux bayésiens
  - 1982 : *Reverend Bayes on inference engines : A distributed hierarchical approach*
    `P(X=vrai)=0.3 et P(Z=faux)=0.2 ...`
    `P(Y=absent)=?`
  - 1988 : *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference.* Morgan Kaufmann

Références

- **Les Réseaux Bayésiens** - P. Naïm, P.H. Wuillemin, Ph. Leray, O. Pourret, A. Becker (Eyrolles)
- **Probabilistic reasoning in Intelligent Systems : Networks of plausible inference** - J. Pearl (Morgan Kaufman)
- **An introduction to Bayesian Networks** - F. Jensen (Springer Verlag)
- **Probabilistic Networks and Expert Systems** - R.G. Cowell & al. (Springer Verlag)
- **Learning Bayesian Networks** - R. Neapolitan (Prenctice Hall)
- **Learning in Graphical Models** - Jordan M.I. ed. (Kluwer)
- **Dynamic Bayesian Networks : Representation, Inference and Learning** - K. Murphy - PhD Thesis, Berkeley

Plan - jour 1

- Représentation de l'incertain
- Rappels de probabilités
- Définition d'un réseau bayésien
- Algorithmes d'inférence
  - Bucket Elimination
  - Message Passing (Pearl)
  - Junction Tree (Jensen)

- MATLAB :
  - Création d'un réseau bayésien
  - Quelques exemples d'inférence

Plan - jour 2

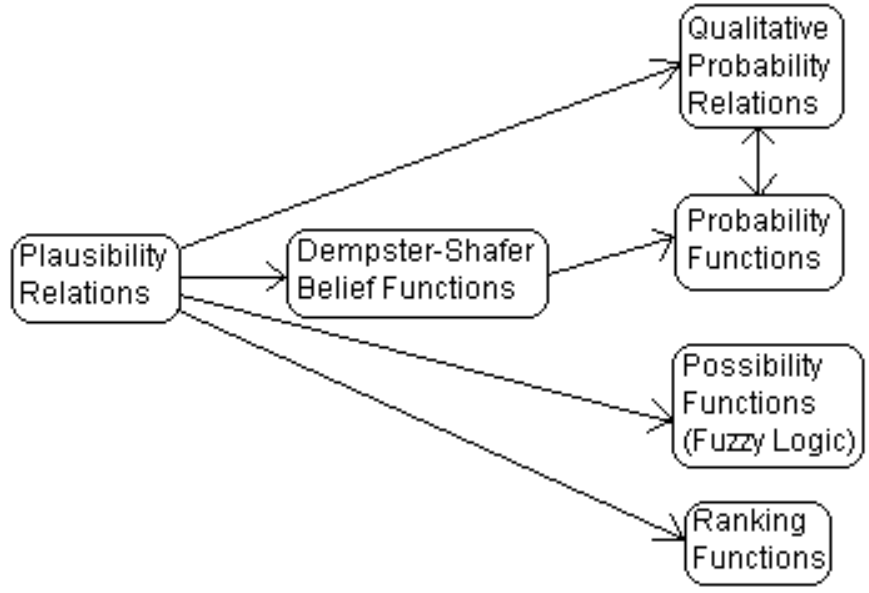
- Apprentissage d'un réseau bayésien
  - Apprentissage des probabilités conditionnelles
    - Expertise / données
    - Données complètes / incomplètes
  - Apprentissage de la structure

- MATLAB : apprentissage d'un RB
- Quelques exemples d'extensions
  - Extension aux variables continues
  - Extension causale
  - Extension temporelle
  - Extension à la décision
- Et les modèles non dirigés ?

Représenter l'incertain

$$Res \in \{A, B, C, D\}$$

- théorie des ensembles : $Res = \{A \; ou \; B\}$
- théorie des probabilités : $P(Res) = [0.7\; 0.3\; 0\; 0]$ (aléatoire)
- théorie des ensembles flous : $Res = faible$ avec $f(faible) = [0.7 \;\; 0.3 \;\; 0 \;\; 0]$ (imprécision)
- théorie de Dempster-Shafer ...



Rappels de probabilités

- Probabilité conditionnelle
  - A et M deux événements
  - information a priori sur A : $P(A)$
  - M s'est produit : $P(M) \neq 0$
  - s'il existe un lien entre A et M, cet événement va modifier notre connaissance sur A
  - information a posteriori : $P(A|M) = \frac{P(A,M)}{P(M)}$

Rappels de probabilités

- Indépendance
  - A et B sont indépendants ssi :

$$P(A, B) = P(A) \times P(B)$$
$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

- Indépendance conditionnelle
  - A et B sont indépendants conditionnellement à C ssi :

$$P(A|B, C) = P(A|C)$$

Rappels de probabilités

$\{M_i\}$ ensemble complet d'événements mutuellement exclusifs

- Marginalisation : $P(A) = \sum_i P(A, M_i)$
- Théorème des probabilités totales :

  Un événement A peut résulter de plusieurs causes M_i.
  Quelle est la probabilité de A connaissant :
  - *les probabilités élémentaires $P(Mi)$ (a priori)*
  - *les probabilités conditionnelles de A pour chaque Mi*

$$P(A) = \sum_i P(A|M_i)P(M_i)$$

mais comment répondre à la question inverse ?

Rappels de probabilités

$\{M_i\}$ ensemble complet d'événements mutuellement exclusifs

- Théorème de Bayes :

  Un événement A s'est produit. Quelle est la probabilité que ce soit la cause M_i qui l'ait produit ?

$$P(M_i|A) = \frac{P(A|M_i) \times P(M_i)}{P(A)}$$

  - $P(Mi|A)$: probabilité a posteriori
  - $P(A)$: constante (pour chaque M_i) cf. th. probas tot.

- Théorème de Bayes généralisé *(Chain rule)*

$$P(A_1 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \dots P(A_n|A_1 \dots A_{n-1})$$

Définition d'un réseau bayésien

- Principe :
  - prendre en compte les indépendances conditionnelles entre les variables pour simplifier la loi jointe donnée par le théorème de Bayes généralisé.

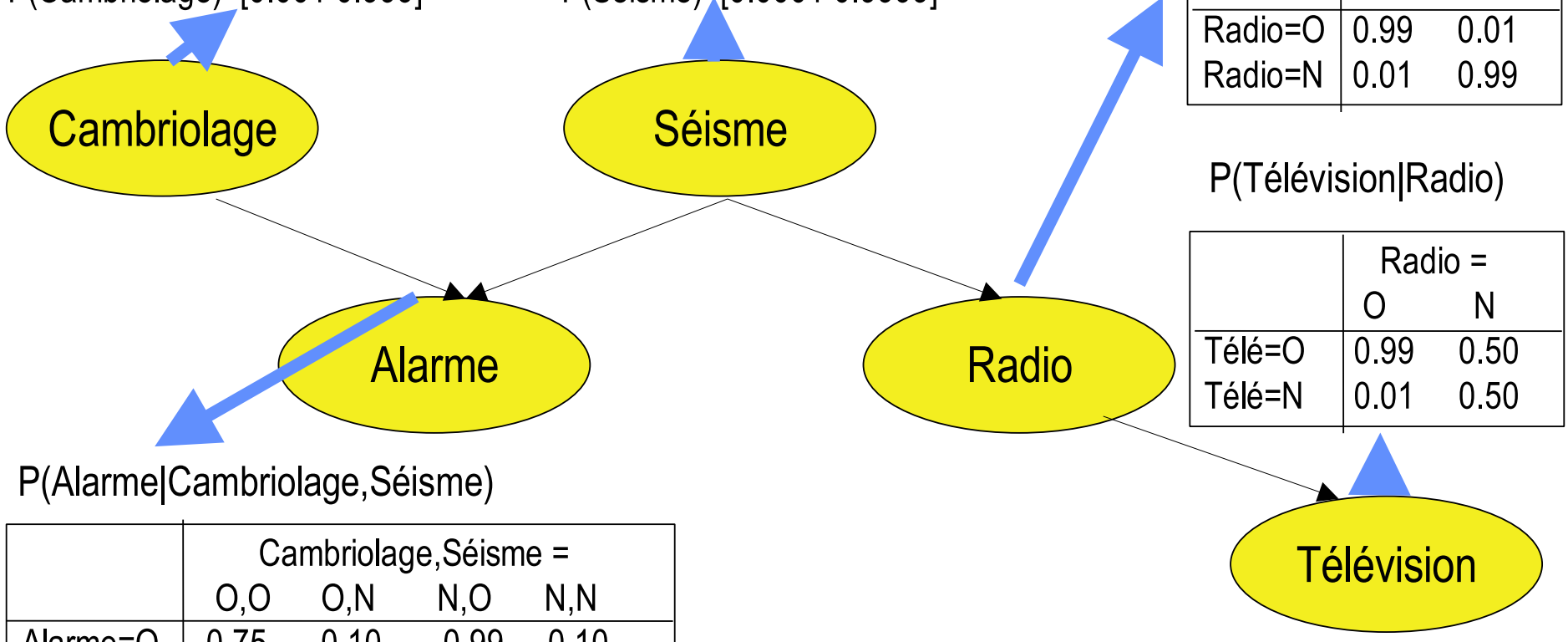
- Un réseau bayésien est défini par
  - la description qualitative des dépendances (ou des indépendances conditionnelles) entre des variables

    graphe dirigé sans circuit (DAG)
  - la description quantitative de ces dépendances

    probabilités conditionnelles (CPD)

Exemple

ordre topologique : C, S, A, R, T (non unique)



P(Radio|Séisme)

	Séisme = O	N
Radio=O	0.99	0.01
Radio=N	0.01	0.99

P(Télévision|Radio)

	Radio = O	N
Télé=O	0.99	0.50
Télé=N	0.01	0.50

P(Alarme|Cambriolage,Séisme)

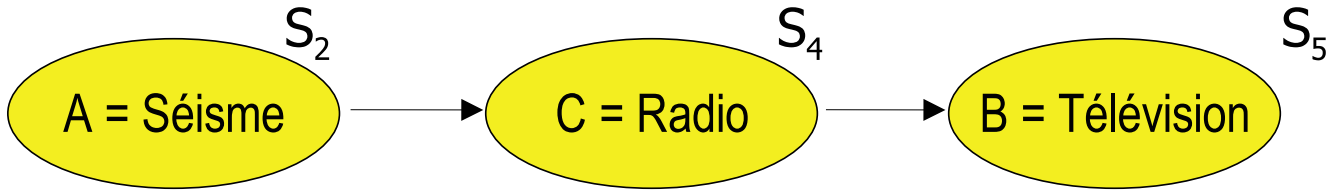
	Cambriolage,Séisme = O,O	O,N	N,O	N,N
Alarme=O	0.75	0.10	0.99	0.10
Alarme=N	0.25	0.90	0.01	0.90

RB et indépendance conditionnelle

- Les RB représentent graphiquement les indépendances conditionnelles

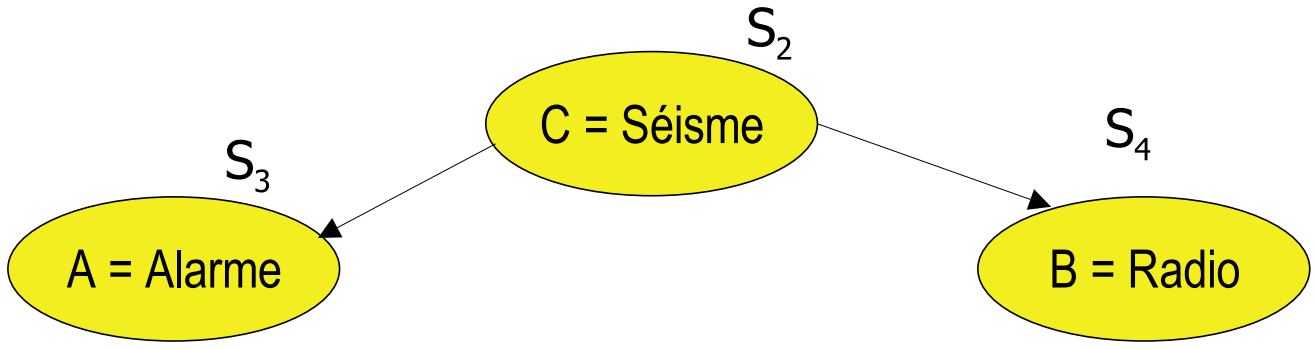
- Exemple sur 3 nœuds
  - 3 types de relations (simples) entre A, B et C :
    - $A \rightarrow C \rightarrow B$: connexion série
    - $A \leftarrow C \rightarrow B$: connexion divergente
    - $A \rightarrow C \leftarrow C$: connexion convergente (V-structure)

Connexion série



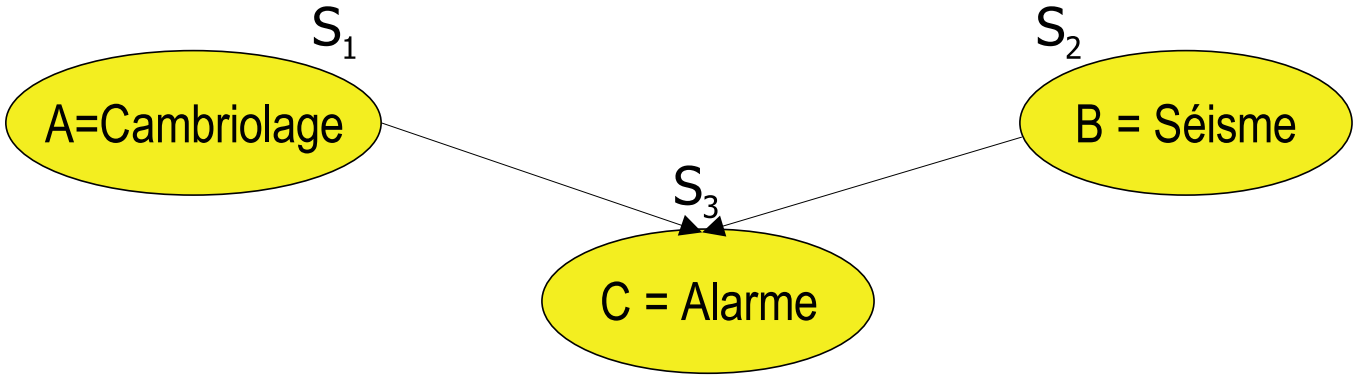
- A et B sont dépendants
- A et B sont indépendants conditionnellement à C
  - si $P(C)$ est connue,
    A n'intervient pas dans le calcul de $P(B)$
  - $P(S_5|S_4, S_2) = P(S_5|S_4) = P(S_5|parents(S_5))$

Connexion divergente



- A et B sont dépendants
- A et B sont indépendants conditionnellement à C
  - si $P(C)$ est connue,
    A n'intervient pas dans le calcul de $P(B)$
  - $P(S_4|S_2, S_3) = P(S_4|S_2) = P(S_4|parents(S_4))$

Connexion convergente – V-structure



- A et B sont indépendants
- A et B sont dépendants conditionnellement à C
  - si $P(C)$ est connue,
    $P(A)$ intervient pas dans le calcul de $P(B)$
  - $P(S_3|S_1, S_2) = P(S_3|parents(S_3))$

Conséquence

- RB = représentation compacte de la loi jointe $P(S)$
- Théorème de Bayes généralisé :

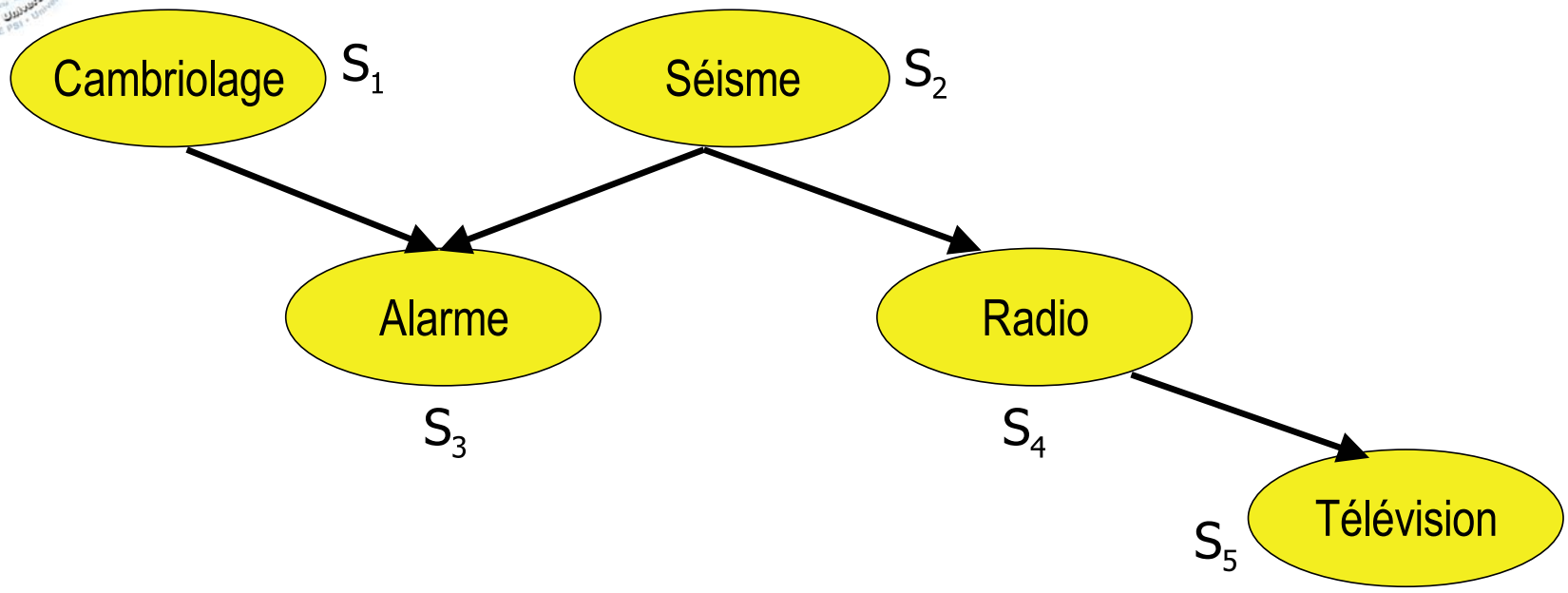
$$P(S) = P(S_1) \times P(S_2|S_1) \times P(S_3|S_1, S_2) \times \cdots \times P(S_n|S_1 \ldots S_{n-1})$$

- Dans un RB, $P(S_i|S_1 \ldots S_{i-1}) = P(S_i|parents(S_i))$ d'où

$$P(S) = \Pi_{i=1}^{n} P(S_i|parents(S_i))$$

- La loi jointe (globale) se décompose en un produit de lois conditionnelles locales

Exemple

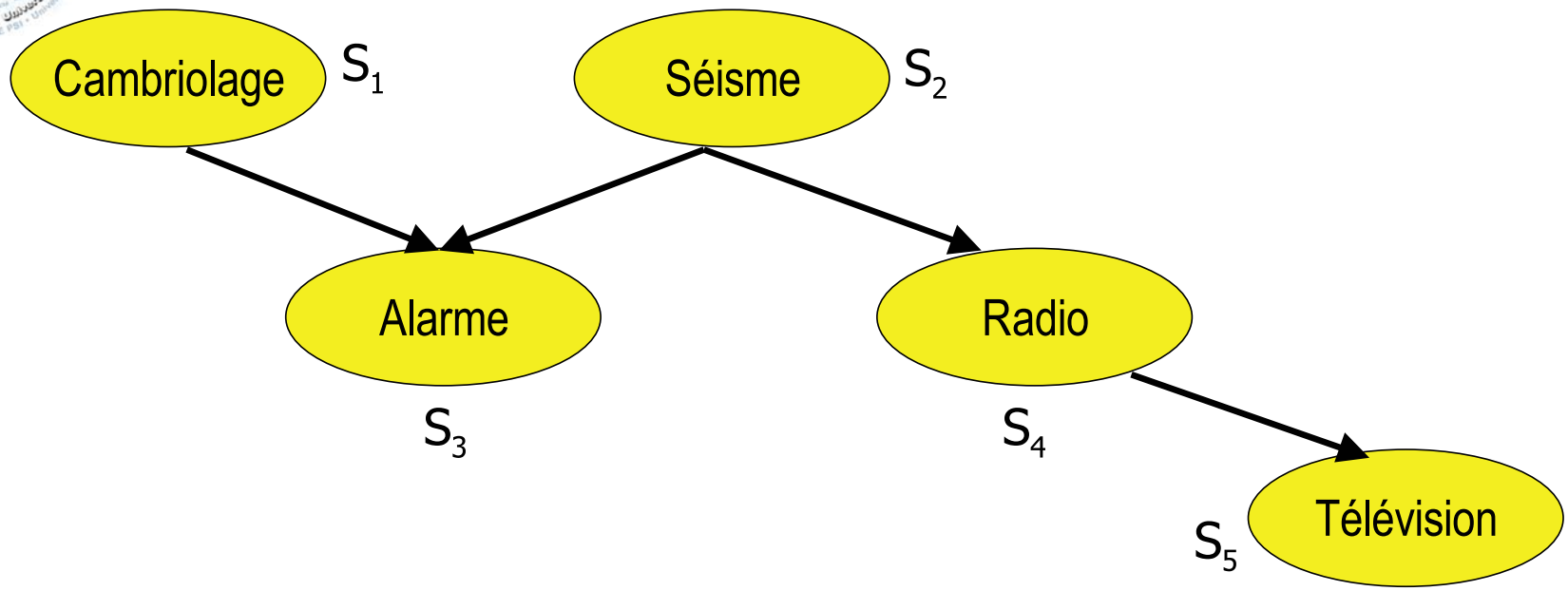


$$P(Cambriolage, Seisme, Alarme, Radio, Tele) =$$
$$P(S_1)P(S_2|S_1)P(S_3|S_1,S_2)P(S_4|S_1,S_2,S_3)P(S_5|S_1,S_2,S_3,S_4)$$
$$P(S_1) \quad P(S_2) \quad P(S_3|S_1,S_2) \quad P(S_4|S_2) \quad P(S_5|S_4)$$

La d-séparation

- Principe
  - déterminer si 2 variables quelconques sont indépendantes conditionnellement à un ensemble de variables instantiées
- Définition
  - deux variables A et B sont d-séparées si pour tous les chemins entre A et B, il existe une variable intermédiaire V différente de A et B telle que
    - la connexion est série ou divergente et V est instancié
    - la connexion est convergente et ni V ni ses descendants ne sont instanciés
- Si A et B ne sont pas d-séparés, ils sont d-connectés

Exemple

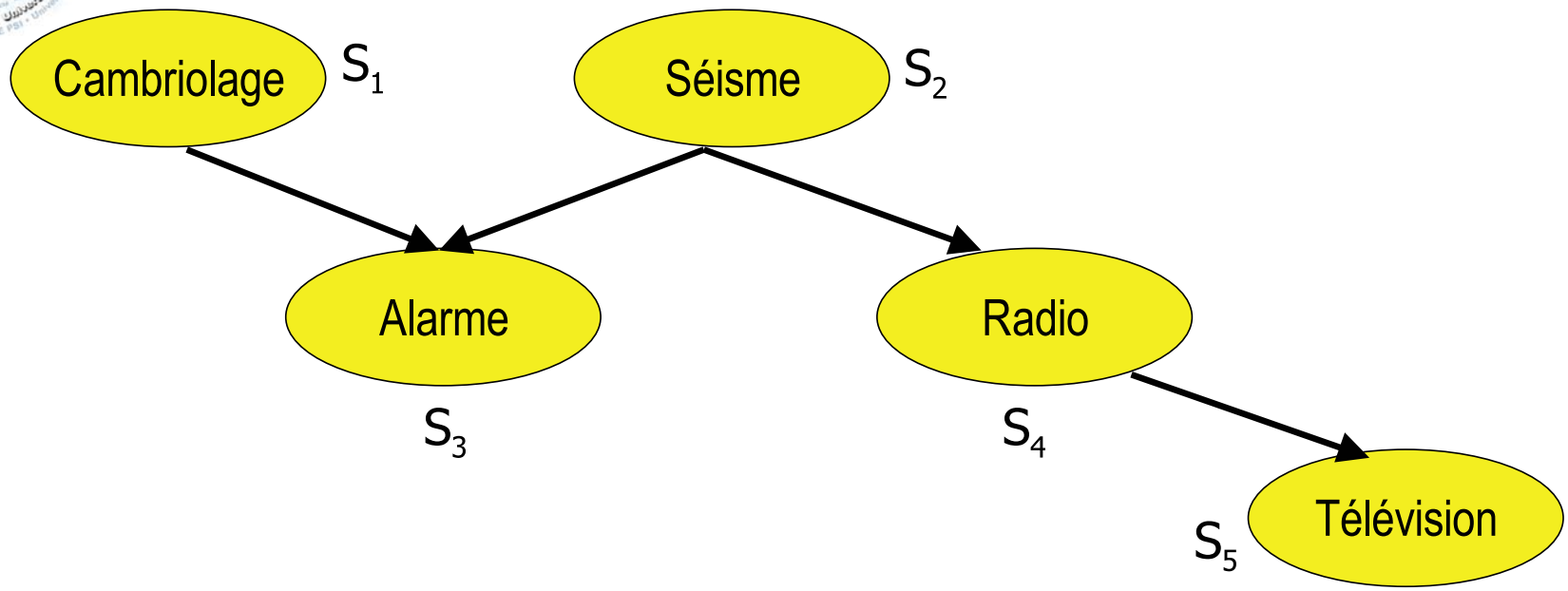


- exemple de d-séparation : $S_1 \ldots S_4$?
  - $V = S_3$ sur le chemin entre S_1 et S_4.
  - la connexion est convergente en V
  - V n'est pas instancié

$\rightarrow$ S_1 et S_4 sont d-séparés

(si S_3 était mesuré, S_1 et S_4 seraient d-connectés)

Exemple



- autre exemple de d-séparation : $S_2 \ldots S_5$?
  - $V = S_4$ sur le chemin entre S_2 et S_5.
  - la connexion est série en V
  - V n'est pas instancié

$\rightarrow$ S_2 et S_5 sont d-connectés

(si S_4 était mesuré, S_2 et S_5 seraient d-séparés)

Plan - jour 1

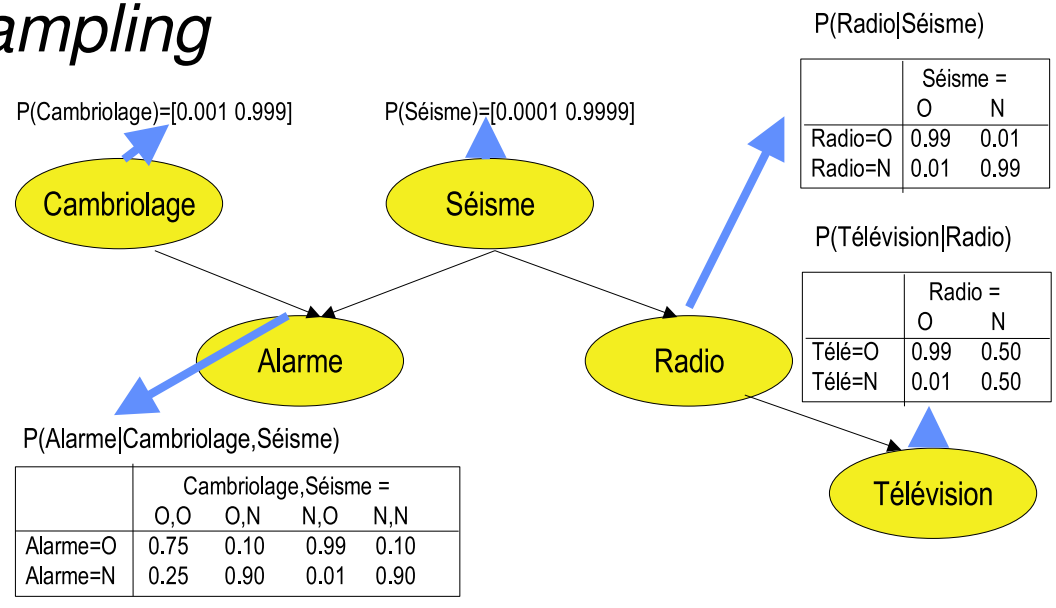
- Représentation de l'incertain
- Rappels de probabilités
- Définition d'un réseau bayésien

- Algorithmes d'inférence
  - Bucket Elimination
  - Message Passing (Pearl)
  - Junction Tree (Jensen)

- MATLAB :
  - Création d'un réseau bayésien
  - Quelques exemples d'inférence

RB = modèle génératif

- RB = représentation compacte de la loi jointe $P(S)$
- Utilisation de méthodes d'échantillonnage pour générer des données qui suivent cette loi
- Exemple : *forward sampling*



- si $rand1 < 0.001$, $C = O$, sinon N
- si $rand2 < 0.0001$, $S = O$, sinon N
- si $rand3 < P(A = O|C = ..., S = ...)$, $A = O$, sinon N
- ...

Inférence

- = calcul de n'importe quelle $P(S_i|S_j = x)$ (NP-complet)

  NB : l'observation $\{S_j = x\}$ est appelée l'évidence

  (soft evidence $S_j = x|_{0.3} \;\; y|_{0.7}$)

- Algorithmes exacts
  - Bucket Elimination
  - Message Passing (Pearl 88) pour les arbres
  - Junction Tree (Jensen 90)
  - Shafer-Shenoy (1990)

Problème = explosion combinatoire de ces méthodes pour des graphes fortement connectés.

- Algorithmes approchés
  - Echantillonnage
  - Méthodes variationnelles

Bucket Elimination

$$P(S_i|S_j = x) = \frac{P(S_i, S_j{=}x)}{P(S_j{=}x)} = \frac{\sum_{\{s_k\}_{k\neq i,j}} P(S_1{=}s_1, \ldots, S_i, S_j{=}x, \ldots S_n{=}s_n)}{\sum_{\{s_k\}_{k\neq j}} P(S_1{=}s_1, \ldots, S_j{=}x, \ldots S_n{=}s_n)}$$

- Principe
  - grâce à la décomposition de $P(S)$, simplifier le calcul de $P(S_i, S_j = x)$
- Exemple
  - évidence $E = \{S_4 = O\}$, on cherche $P(S_2|E)$

$$P(S, E) = P(S_1)P(S_2)P(S_3|S_1S_2)P(S_4 = O|S_2)P(S_5|S_4 = O)$$

$$P(S_2, E) = \sum_{S_1, S_3, S_5} P(S, E)$$

  - et si on choisit l'ordre des S_i pour la marginalisation ?

Bucket Elimination

- Commençons par S_5

$$\sum_{S_5} P(S_1, S_2, S_3, S_4 = O, S_5) = P(S_1)P(S_2)P(S_3|S_1, S_2)\ldots$$

$$\ldots \quad P(S_4 = O|S_2) \sum_{S_5} P(S_5|S_4 = O)$$

- Cette dernière somme vaut 1 ! On a éliminé S_5

$$P(S_1, S_2, S_3, S_4 = O) = P(S_1)P(S_2)P(S_3|S_1, S_2)P(S_4 = O|S_2)$$

- Au tour de S_1

Bucket Elimination

$$\sum_{S_1} P(S_1, S_2, S_3, S_4 = O) \quad = \quad P(S_2)P(S_4 = O|S_2) \ldots$$

$$\ldots \quad \sum_{S_1} P(S_1)P(S_3|S_1, S_2)$$

- Cette dernière somme nous rend une table dépendant de S_2 et S_3 : $T(S_2, S_3)$

$$P(S_2, S_3, S_4 = O) \quad = \quad P(S_2)P(S_4 = O|S_2)T(S_2, S_3)$$

- Idem avec S_3 pour obtenir $P(S_2, S_4 = O)$

Marginalisation = série de produits locaux de matrices et de marginalisations locales

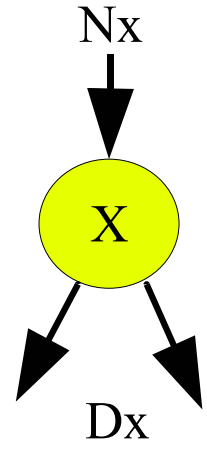
Bucket Elimination

- Comment bien choisir l'ordre de marginalisation ?
  - problème lui aussi N-P complet :-)
  - quelques heuristiques efficaces
    - minimum deficiency search (Bertele and Brioschi 1972, Kjærulff 1990)
    - maximum cardinality search (Tarjan and Yannakakis 1984)

Message Passing (Pearl 1988)

- Chaque nœud envoie des messages à ses voisins
- L'algorithme ne marche que dans le cas des arbres ((mais est généralisable au cas des poly-arbres)

- E = ensemble de variables instanciées.
  $E = N_x \cup D_x$



- 2 types de messages λ et π serviront à calculer
  - $\lambda(X) \propto P(D_x|X)$
  - $\pi(X) \propto P(X|N_x)$

- et ensuite on peut montrer que

$$P(X|E = e) \propto \lambda(X)\pi(X)$$

Message Passing

- Les messages λ
  - Pour chaque enfant Y de X,

$$\lambda_Y(X = x) = \sum_y P(Y = y|X = x)\lambda(Y = y)$$

- Comment calculer λ en chaque nœud ?
  - Si X instancié, $\lambda(X) = [001 \ldots 0]$
    (la position du 1 correspond à la valeur donnée à X
  - sinon
    - si X est une feuille, $\lambda(X) = [1 \ldots 1]$
    - sinon

$$\lambda(X = x) = \Pi_{Y \in Enf(X)} \lambda_Y(X = x)$$

Message Passing

- Les messages π
  - Pour Z l'unique parent de X,

$$\pi_X(Z = z) = \pi(Z = z)\Pi_{U \in Enf(Z) \setminus \{X\}} \lambda_U(Z = z)$$

- Comment calculer π en chaque nœud ?
  - Si X instancié, $\lambda(X) = [001 \ldots 0]$
    (la position du 1 correspond à la valeur donnée à X
  - sinon
    - si X est la racine, $\pi(X) = P(X)$
    - sinon

$$\pi(X = x) = \sum_z P(X = x | Z = z) \pi_X(Z = z)$$

Junction Tree (Jensen 1990)

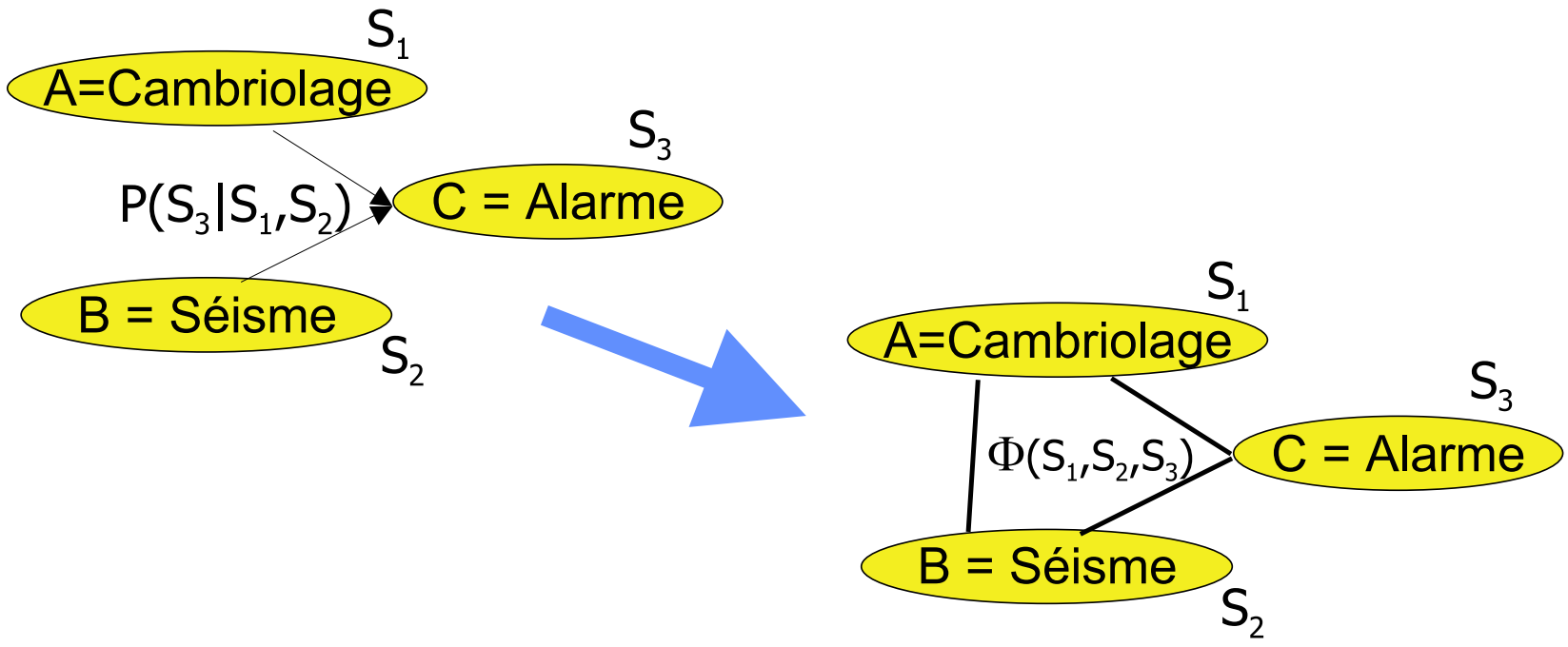
- Message Passing ne s'applique bien qu'aux arbres
- Besoin d'un algorithme plus général

- Principe
  - Transformer le graphe en un arbre (non orienté)...
  - Arbre = arbre de jonction des cliques maximales du graphe moralisé et triangulé

- Moralisation = marier les parents et "désorienter" le graphe
- Triangulation = éviter les cycles dans le graphe non orienté.

Junction Tree

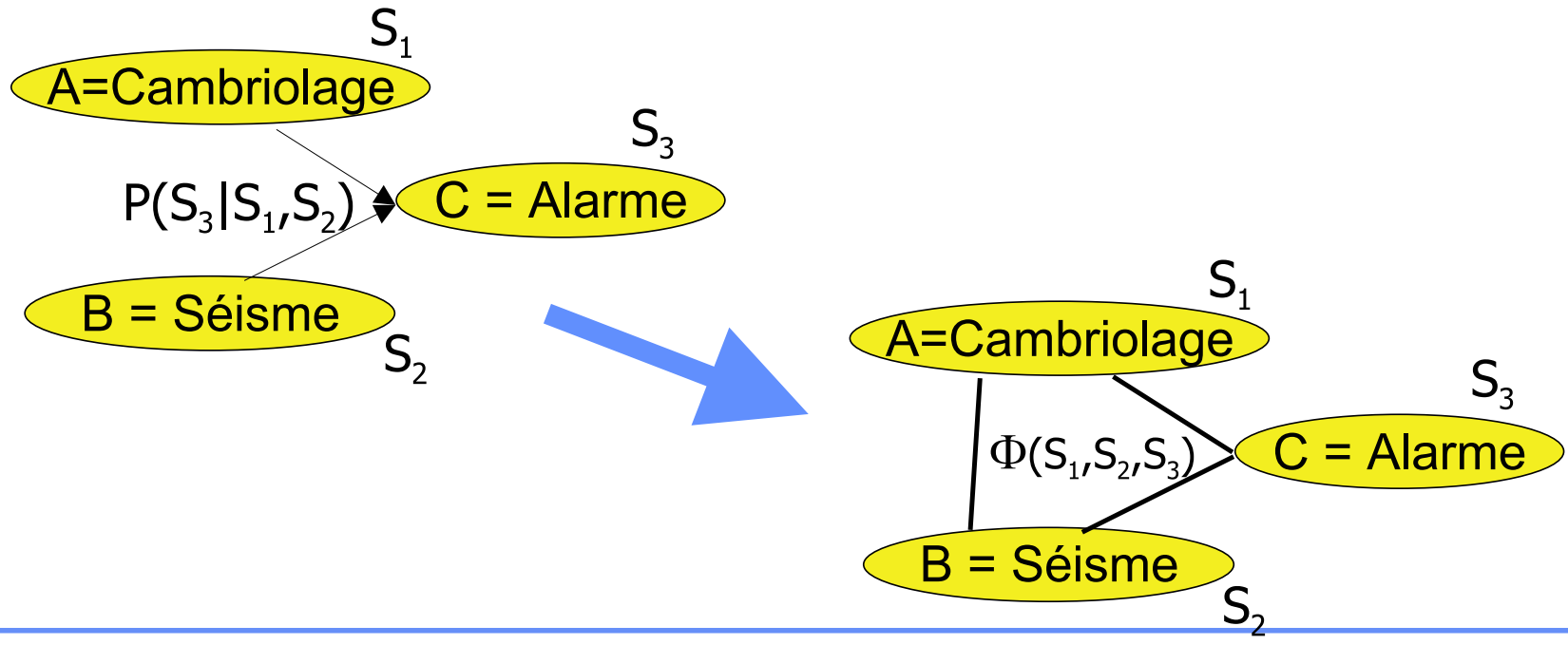
- Moralisation : marier les parents de chaque nœud



Junction Tree

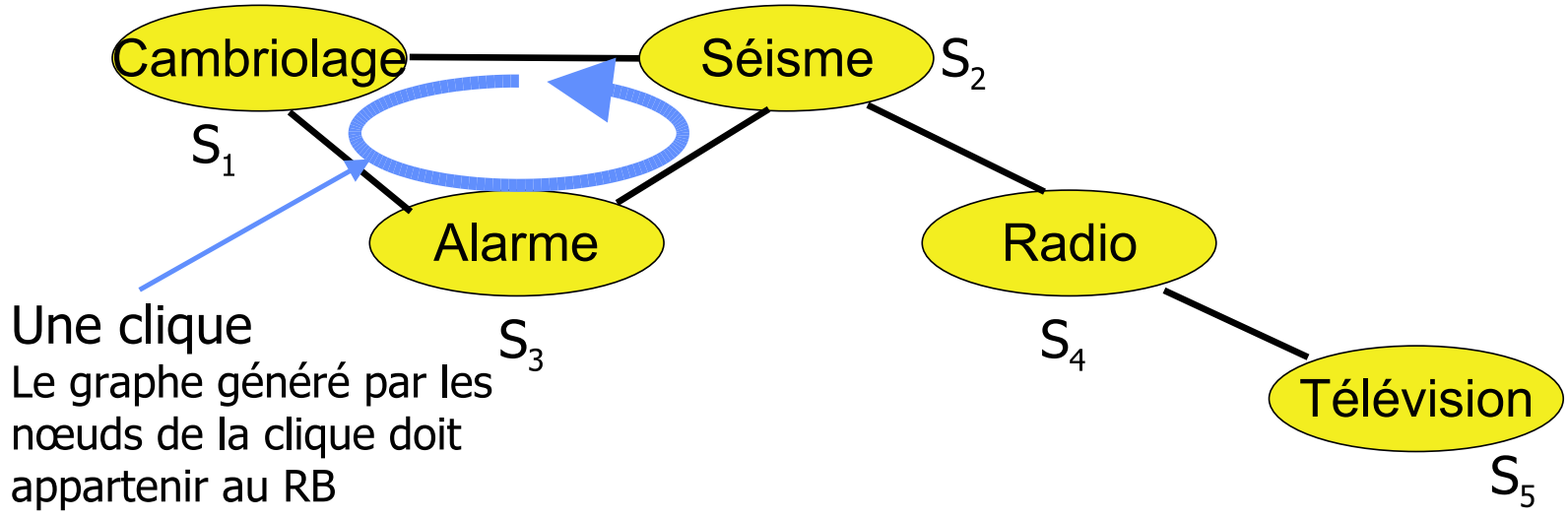
- Triangulation : tout cycle de longueur au moins 4 doit contenir une corde (arête reliant deux sommets non consécutifs sur le cycle)
- (= aucun sous-graphe cyclique de longueur > 3).

- Triangulation optimale pour des graphes non-dirigés = NP-difficile (comment choisir les meilleures cordes ?)



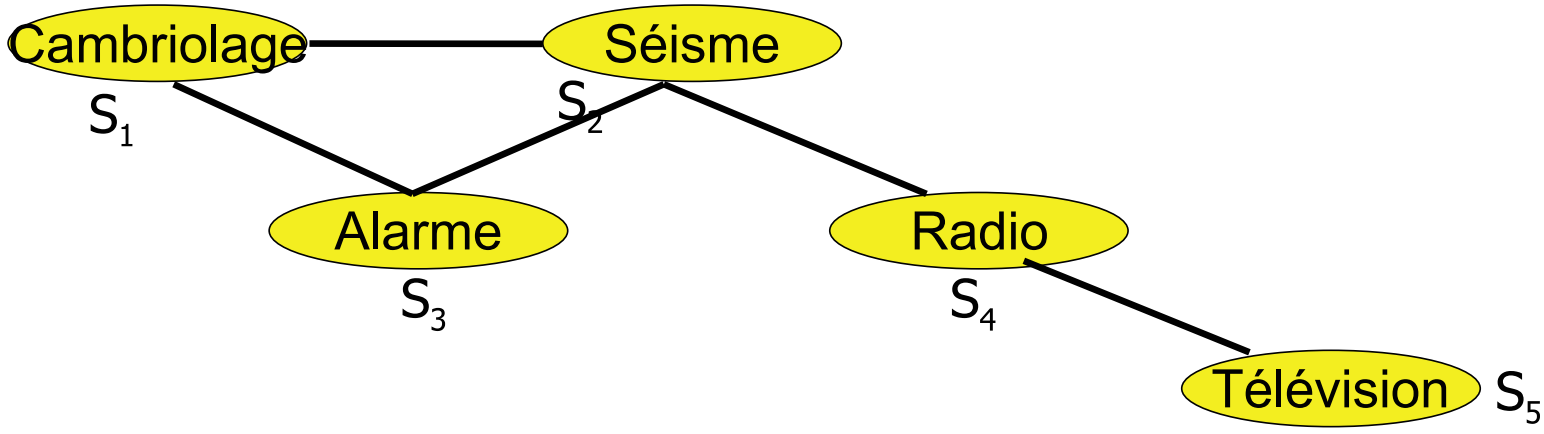
Junction Tree

- Clique = sous-graphe du RB dont les nœuds sont complétement connectés
- Clique maximale = l'ajout d'un autre nœud à cette clique ne donne pas une clique



Junction Tree

- Théorème : Si le graphe est moralisé et triangulé, alors les cliques peuvent être organisées en un arbre de jonction



$$P(S) = \Phi(S1, S2, S3)\Phi(S2, S4)\Phi(S4, S5)$$

- L'inférence se fait au niveau des Φ

Inférence abductive

- Autres types d'inférence
  - MPE (Most Probable Explanation) - la configuration de TOUTES les variables la plus probable sachant l'évidence
  - MAP (Maximum A Posteriori) - la configuration d'un ENSEMBLE de variables la plus probable sachant l'évidence
- En général $MPE(S|E) \neq \{argmax\ P(S_i|E)\}_i$
- idem $MAP(S_i, S_j|E) \neq MPE(S|E)|_{S_i,S_j}$
- Adaptation des algorithmes d'inférence classiques

Plan - jour 1

- Représentation de l'incertain
- Rappels de probabilités
- Définition d'un réseau bayésien
- Algorithmes d'inférence
  - Bucket Elimination
  - Message Passing (Pearl)
  - Junction Tree (Jensen)

- MATLAB :
  - Création d'un réseau bayésien
  - Quelques exemples d'inférence