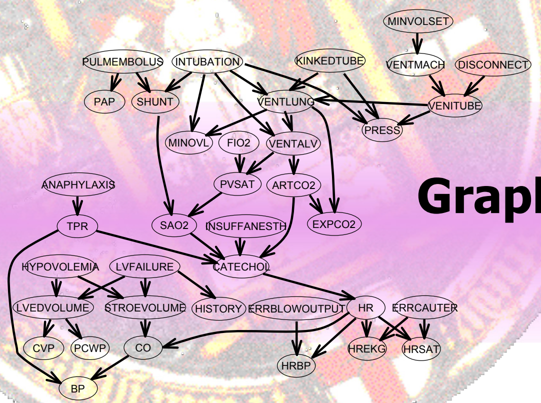
Mainly based on F. V. Jensen, „Bayesian Networks and Decision Graphs“, Springer-Verlag New York, 2001.

Advanced I WS 06/07



Graphical Models - Modeling -

Wolfram Burgard, Luc De Raedt, Kristian Kersting, Bernhard Nebel

Albert-Ludwigs University Freiburg, Germany

Outline

- Introduction
- Reminder: Probability theory
- Basics of Bayesian Networks
- Modeling Bayesian networks
- Inference
- Excourse: Markov Networks
- Learning Bayesian networks
- Relational Models

(Discrete) Random Variables and Probability

A ***random variable*** X describes a value that is the result of a process which can not be determined in advance: the rank of the card we draw from a deck

The ***sample space*** S of a random variable is the set of all possible values of the variable: Ace, King, Queen, Jack, …

An ***event*** is a subset of S: {Ass, King}, {Queen}, …

The **probability function** P(X=x), short P(x),defines the probability that X yields a certain value x.

(Discrete) Random Variables and Probability

The **world** Ω is described as a set of random variables and divided it into elementary events or ***states***

$$s_1, s_2, \ldots$$

States are **mutually exclusive**

$$P(s_i \wedge s_j) = P(s_i, s_j) = 0 \text{ if } i \neq j$$

A **probability space** has the following properties:

$$0 \leq P(X) \leq 1 \qquad P(\mathit{false}) = 0 \qquad P(\mathit{true}) = 1$$

$$\sum_x P(X = x) = \sum_x P(x) = 1$$

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Conditional probability and Bayes' Rule

If it is known that a certain event occurred this may change the probability that another event will occur

P(x|y) denotes the ***conditional probability*** that x will happen given that y has happened.

Bayes' rule states that:

$$P(A|B) = \frac{P(A,B)}{P(B)} = P(B|A)\frac{P(A)}{P(B)}$$
if $P(B) > 0$

Fundamental rule

- The fundamental rule (sometimes called ‚chain rule‘) for probability calculus is

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_{i-1}, \ldots, X_1)$$

$$P(X_1, \ldots, X_n|E) = \prod_{i=1}^{n} P(X_i|X_{i-1}, \ldots, X_1, E)$$

Conditional probability and Bayes' Rule

The ***complete probability formula*** states that

$$P(A) = \sum_b P(A, B=b) = \sum_b P(A|B=b)P(B=b)$$

Note:

$$P(A) = \alpha \cdot P(A|B) + (1-\alpha) \cdot P(A|\neg B)$$

mirrors our intuition that the unconditional probability of an event is somewhere between its conditional probability based on two opposing assumptions.

Joint Probability Distribution

- „truth table“ of set $X_1, \ldots, X_n$ of random variables

X_1	X_2	X_3	X_1, X_2, X_3
true	1	green	0.001
true	1	blue	0.021
true	2	green	0.134
true	2	blue	0.042
...	...	...	...
false	2	blue	0.2

- Any probability we are interested in can be computed from it

Bayes´rule and Inference

Bayes' is fundamental in learning when viewed in terms of evidence and hypothesis:

posterior probability — **likelihood** — **prior probability**

$$P(H|e) = P(e|H) \cdot \frac{P(H)}{P(e)} = \frac{P(e|H) \cdot P(H)}{P(e|H) \cdot P(H) + P(e|\neg H)}$$

This allows us to update our belief in a hypothesis based on prior belief and in response to evidence.

Independence among random variables

Two random variables X and Y are **independent** if knowledge about X does not change the uncertainty about Y and vice versa.

Formally for a probability distribution P:

$$\begin{aligned}\mathrm{Ind}(X;Y) &\equiv P(X|Y) = P(X)\\ &\equiv P(Y|X) = P(Y)\end{aligned}$$

Independence among random variables

$$\textbf{If } \mathrm{Ind}(X;Y) \textbf{ then } P(X,Y) = P(X) \cdot P(Y)$$

$$\textbf{If } \mathrm{Ind}(X;\{Y,Z\}) \textbf{ then } P(X,Y,Z) = P(X) \cdot P(Y,Z)$$

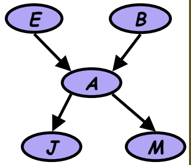
$$\textbf{If } \mathrm{Ind}(X;\{Y,Z\}) \textbf{ then } \mathrm{Ind}(X;Y) \wedge \mathrm{Ind}(X;Z)$$

but not the other way around

$$\textbf{If } \mathrm{Ind}(X;Y|Z\}) \textbf{ then}$$
$$P(X,Y|Z) = P(X|Z) \cdot P(Y|Z) \text{ and } P(X|Y,Z) = P(X|Z)$$

Outline

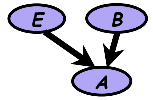
Bayesian Networks



1. Finite, acyclic graph
2. Nodes: (discrete) random variables
3. Edges: direct influences
4. Associated with each node: a table representing a conditional probability distribution (CPD), quantifying the effect the parents have on the node

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i |\, \mathrm{pa}(X_i))$$

Associated CPDs

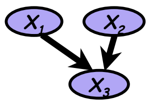


- naive representation
  - tables
- other representations
  - decision trees
  - rules
  - neural networks
  - support vector machines
  - ...

E	B	P(A | E,B)	
e	b	.9	.1
e	b	.7	.3
e	b	.8	.2
e	b	.99	.01

Bayesian Networks

$P(X_1)$
(0.2, 0.8)



$P(X_2)$
(0.6, 0.4)

X_1	X_2	$P(X_3 \| X_2, X_1)$
true	1	(0.2,0.8)
true	2	(0.5,0.5)
false	1	(0.23,0.77)
false	2	(0.53,0.47)

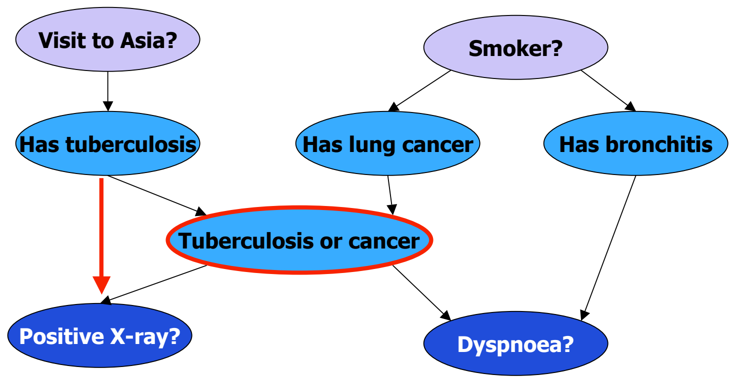
$$\begin{aligned} P(X_1 = \mathit{true}, X_2 = 1, X_3 = \mathit{false}) \\ = P(X_1 = \mathit{true}) \cdot P(X_2 = 1) \\ \cdot P(X_3 = \mathit{false} | X_2 = 1, X_1 = \mathit{true}) \\ = 0.2 \cdot 0.6 \cdot 0.8 = 0.96 \end{aligned}$$

Conditional Independence I

Each node / random variable is (conditionally) independent of all its non-descendants given a joint state of its parents.

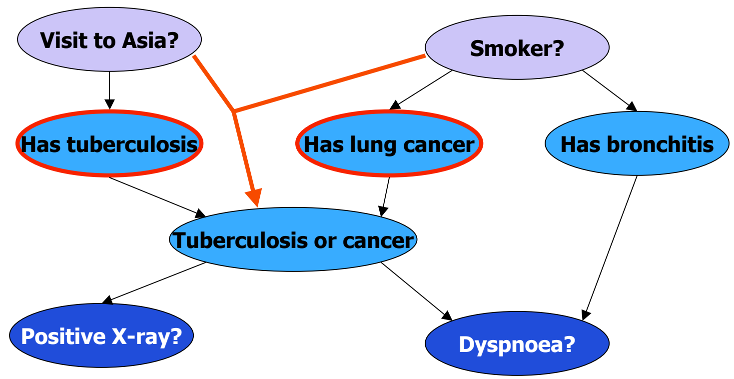
Example

$$\mathrm{Ind}(\mathit{Positive\ X\text{-}ray}; \mathit{Has\ tuberculosis} \mid \mathit{Tuberculosis\ or\ cancer})$$



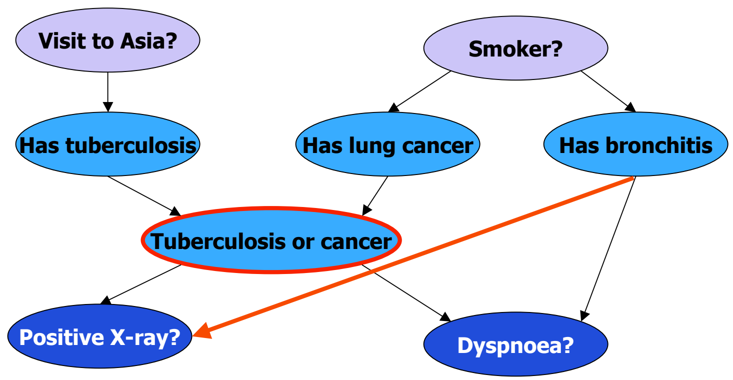
Example

$$\mathrm{Ind}(\mathit{Tuberculosis\ or\ cancer}; \{\mathit{Visit\ to\ Asia}, \mathit{Smoker}\} \mid \{\mathit{Has\ tuberculosis}, \mathit{Has\ lung\ cancer}\})$$



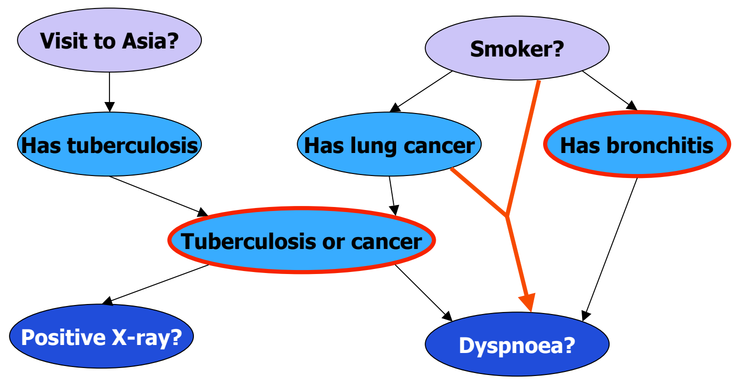
Example

$$\mathrm{Ind}(\mathit{Positive\ X\text{-}ray}; \mathit{Has\ bronchitis} \mid \mathit{Tuberculosis\ or\ cancer})$$

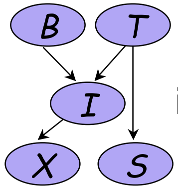


Example

$$\mathrm{Ind}(\mathit{Dyspnoea}; \{\mathit{Has\ lung\ cancer}, \mathit{Smoker}\} \mid \{\mathit{Has\ tuberculosis\ or\ cancer}, \mathit{Has\ bronchitis}\})$$



Summary Semantics



conditional independencies in BN structure **+** local probability models **=** full joint distribution over domain

$$P(\neg B, T, I, X, \neg S) = P(\neg B) \cdot P(T) \cdot P(I|\neg B, T) \cdot P(X|I) \cdot P(\neg S|T)$$

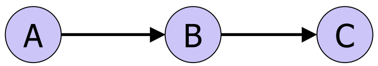
Compact & natural representation:
nodes have ≤ *k* parents
⇒ ***2^k n*** instead of ***2^n*** params

Operations

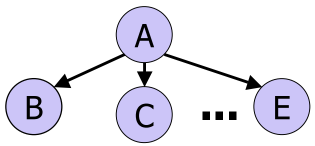
- Inference: $P(Dyspnoea, Smoker|E)$
- Most probable explanation:
  - Most probable configuration of $Dyspnoea, Smoker$ given the evidence E
- Data Conflict (coherence of evidence)
  - positive conf(e) indicates a possible conflict $conf(E) = \log\left(\frac{P(e_1) \cdot P(e_2) \cdot \ldots \cdot P(e_n)}{P(E)}\right)$
- Sensitivity analysis
  - Which evidence is in favour of/against/irrelevant for H
  - Which evidence discriminates H from H´

Conditional Independence II

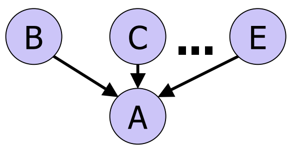
- One could graphically read off other conditional independencies based on
  - Serial connections



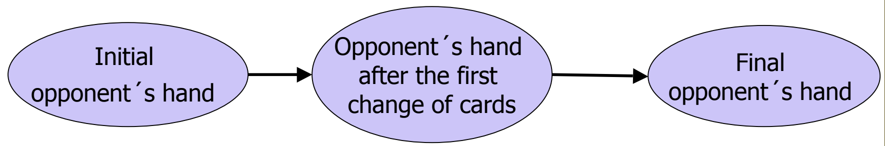
  - Diverging connections



  - Converging connections



Serial Connections - Example

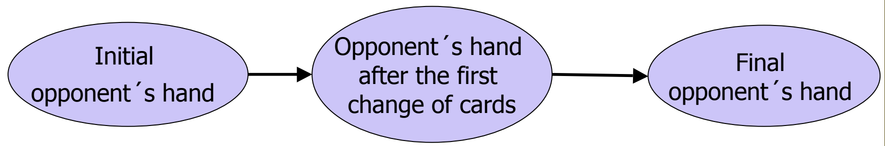


- Poker: If we do **not** know the opponent´s hand after the first change of cards, **knowing** her initial hand will tell us more about her final hand.

Serial Connections - Example

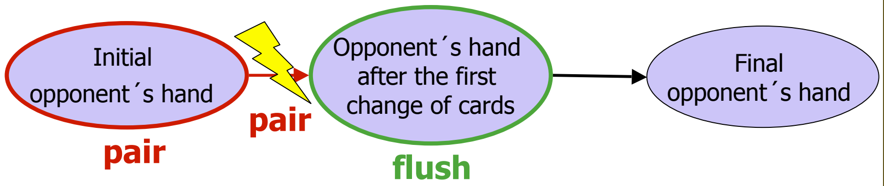
- Poker: If we do **not** know the opponent´s hand after the first change of cards, **knowing** her initial hand will tell us more about her final hand.

Serial Connections - Example



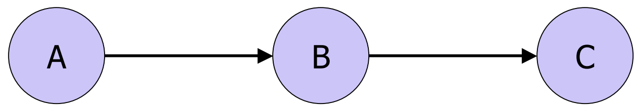
- If we **know** the opponent´s hand after the first change of cards, knowing her initial hand will tell us **nothing** about her final hand.

Serial Connections - Example



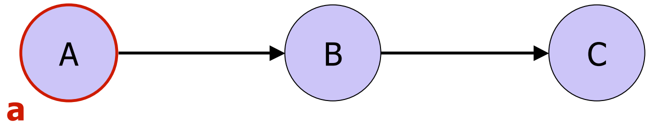
- If we **know** the opponent´s hand after the first change of cards, knowing her initial hand will tell us **nothing** about her final hand.

Serial Connections



- A has influence on B, and B has influence on C

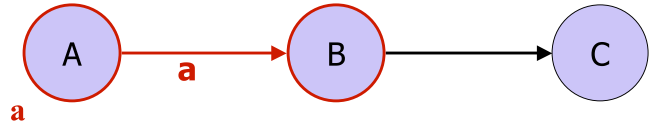
Serial Connections



- A has influence on B, and B has influence on C

Advanced I WS 06/07

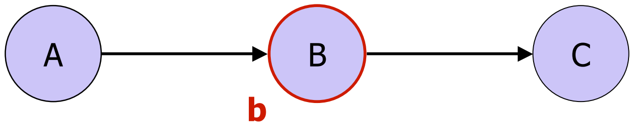
Serial Connections



- A has influence on B, and B has influence on C

Advanced I WS 06/07

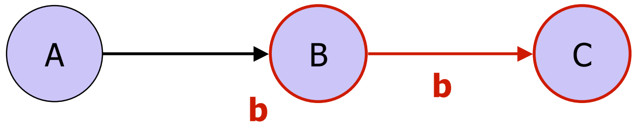
Serial Connections



- A has influence on B, and B has influence on C

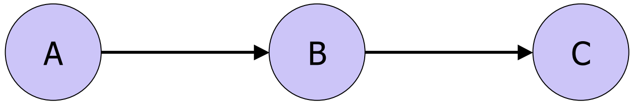
Advanced I WS 06/07

Serial Connections



- A has influence on B, and B has influence on C

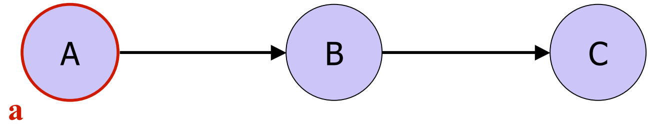
Serial Connections



- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa

Advanced I WS 06/07

Serial Connections

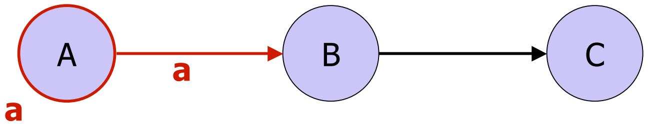


- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

Serial Connections

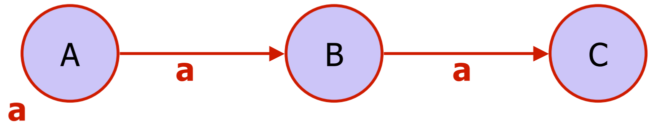


- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

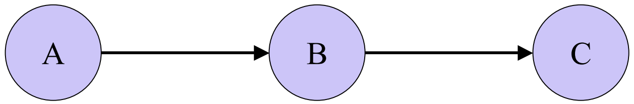
Serial Connections



- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa

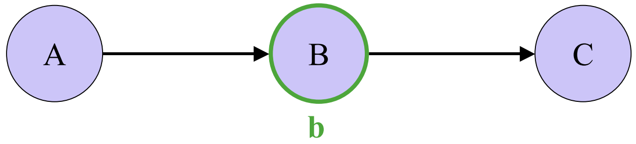
Bayesian Networks - Bayesian Networks

Serial Connections



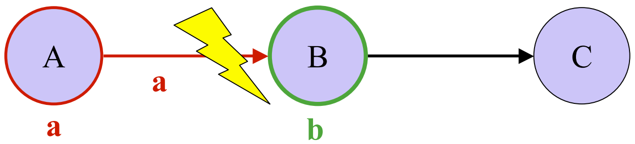
- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa
- **IF** the state of B is known, **THEN** the channel through B is blocked, A and B become independent

Serial Connections



- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa
- **IF** the state of B is known, **THEN** the channel through B is blocked, A and B become independent

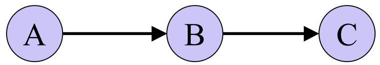
Serial Connections



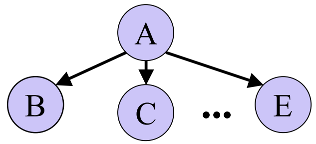
- A has influence on B, and B has influence on C
- Evidence on A will influence the certainty of C (through B) and vice versa
- **IF** the state of B is known, **THEN** the channel through B is blocked, A and B become independent

Conditional Independence II

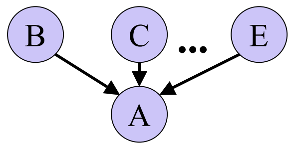
- One could graphically read off other conditional independencies based on
  - Serial connections



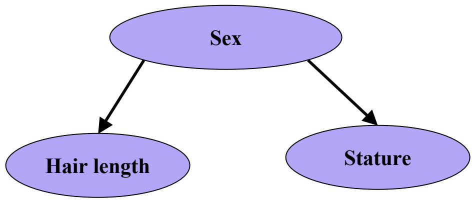
  - **Diverging connections**



  - Converging connections

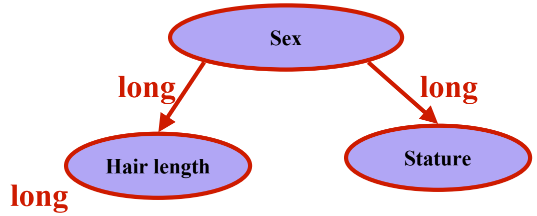


Diverging Connections - Example



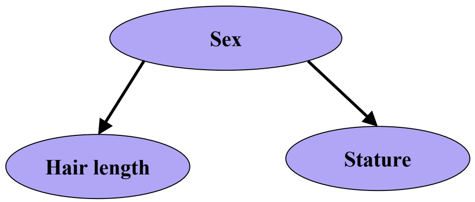
- If we do **not** know the sex of a person, seeing the length of her hair **will tell** us more about the sex, and this in turn will focus our belief on her stature.

Diverging Connections - Example



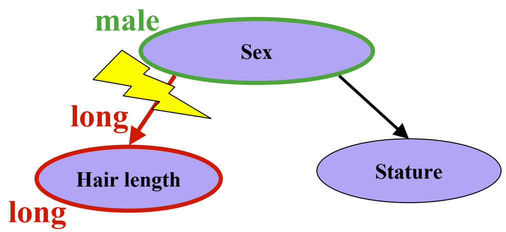
- If we do **not** know the sex of a person, seeing the length of her hair **will tell** us more about the sex, and this in turn will focus our belief on her stature.

Diverging Connections - Example



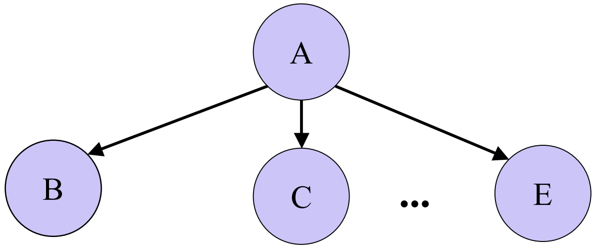
- If we **know** that the person is a man, then the length of hair gives us **no extra clue** on his stature.

Diverging Connections - Example



- If we **know** that the person is a man, then the length of hair gives us **no extra clue** on his stature.

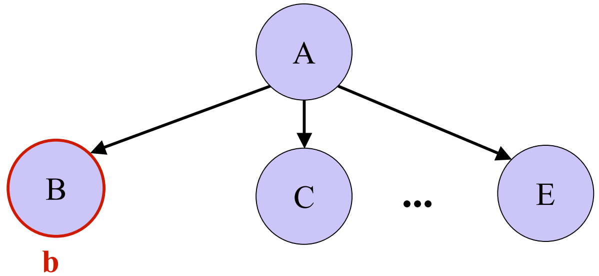
Diverging Connections



- Influence can pass between all the children of A **unless** the state of A is known

Advanced I WS 06/07

Diverging Connections

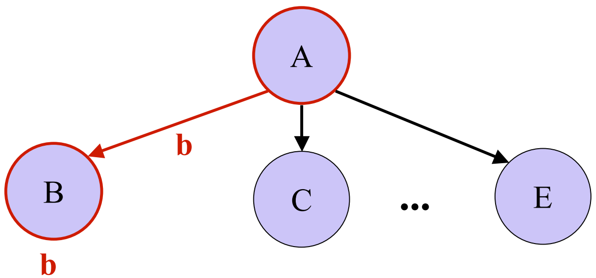


- Influence can pass between all the children of A **unless** the state of A is known

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

Diverging Connections

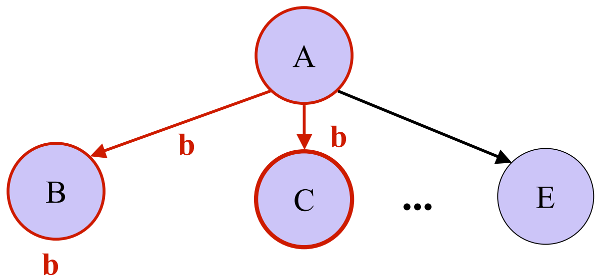


- Influence can pass between all the children of A **unless** the state of A is known

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

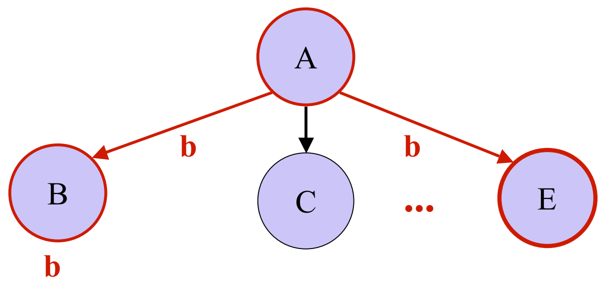
Diverging Connections



- Influence can pass between all the children of A **unless** the state of A is known

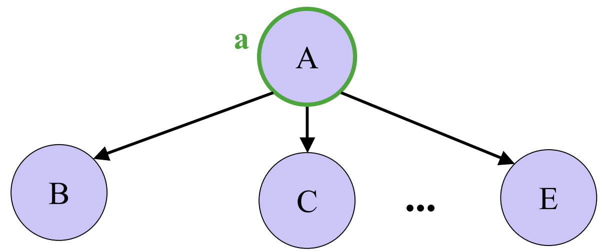
Bayesian Networks - Bayesian Networks

Diverging Connections



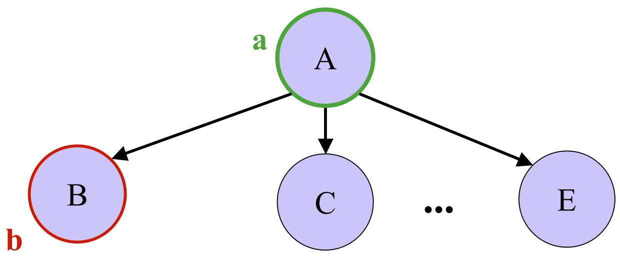
- Influence can pass between all the children of A **unless** the state of A is known

Diverging Connections



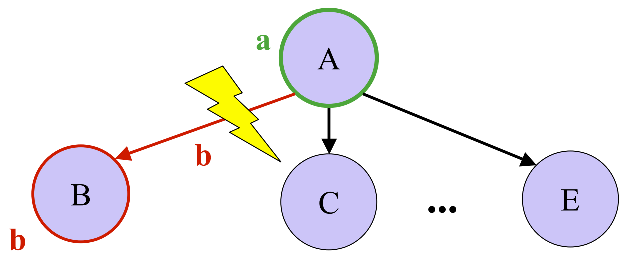
- Influence can pass between all the children of A **unless** the state of A is known

Diverging Connections



- Influence can pass between all the children of A **unless** the state of A is known

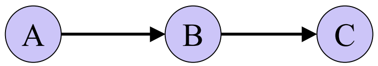
Diverging Connections



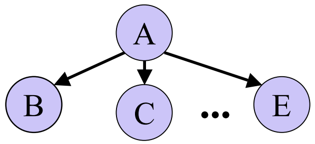
- Influence can pass between all the children of A **unless** the state of A is known

Conditional Independence II

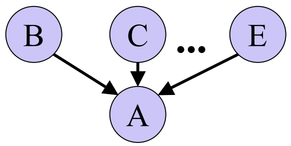
- One could graphically read off other conditional independencies based on
  - Serial connections



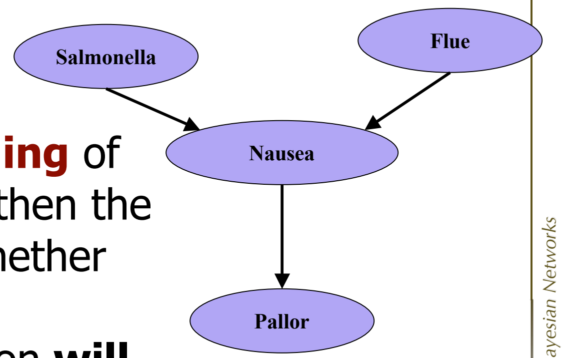
  - Diverging connections



  - **Converging connections**

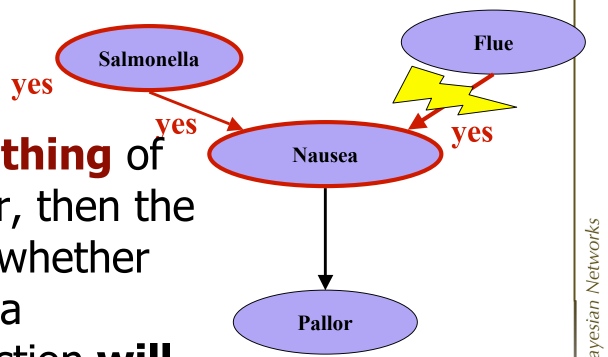


Converging Connections - Example



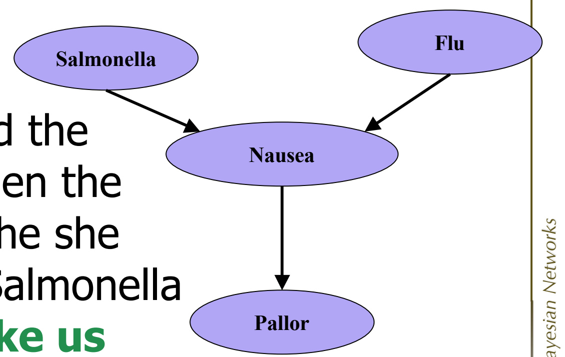
- If we **know nothing** of nausea or pallor, then the information on whether the person has a Salmonella infection **will tell us anything** about flu.

Converging Connections - Example



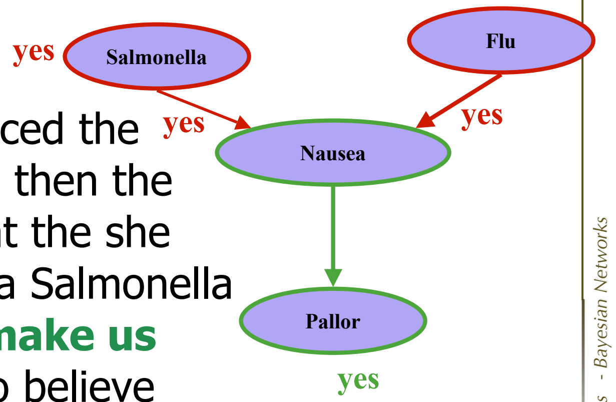
- If we **know nothing** of nausea or pallor, then the information on whether the person has a Salmonella infection **will tell us anything** about flu.

Converging Connections - Example



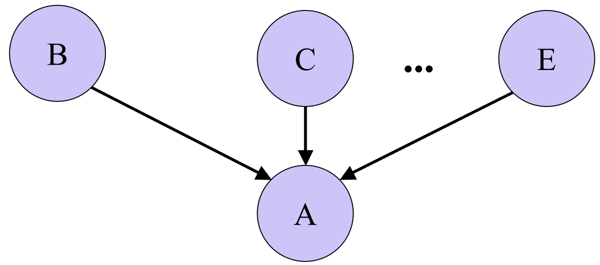
- If we have noticed the person is **pale**, then the information that the she does not have a Salmonella infection **will make us more ready** to believe that she has the flu.

Converging Connections - Example



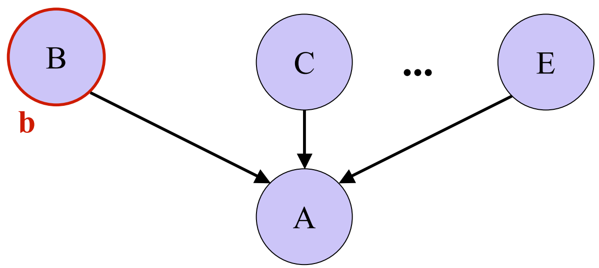
- If we have noticed the person is **pale**, then the information that the she does not have a Salmonella infection **will make us more ready** to believe that she has the flu.

Converging Connections



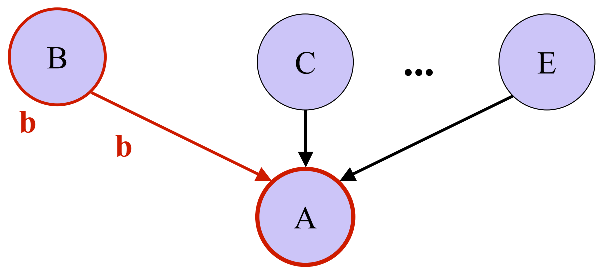
- Evidence of one of A´s parents has **no** influence on the certainty of the others

Converging Connections



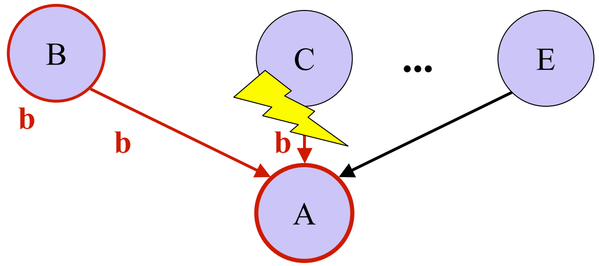
- Evidence of one of A´s parents has **no** influence on the certainty of the others

Converging Connections



- Evidence of one of A´s parents has **no** influence on the certainty of the others

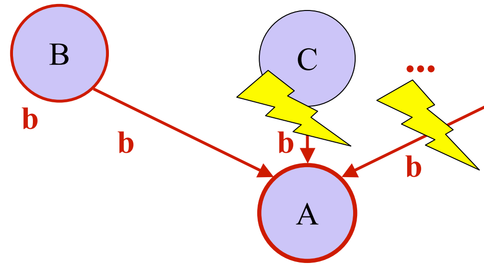
Converging Connections



- Evidence of one of A´s parents has **no** influence on the certainty of the others

Advanced I WS 06/07

Converging Connections

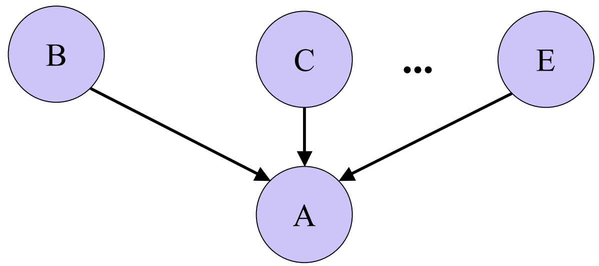


- Evidence of one of A´s parents has **no** influence on the certainty of the others

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

Converging Connections

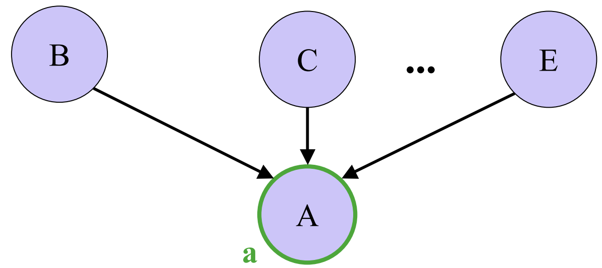


- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

Bayesian Networks - Bayesian Networks

Advanced I WS 06/07

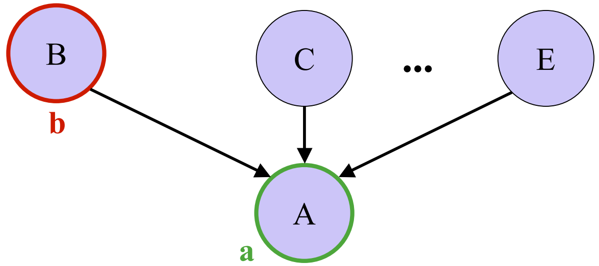
Converging Connections



- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

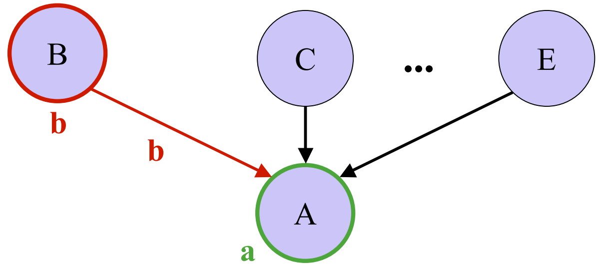
Bayesian Networks - Bayesian Networks

Converging Connections



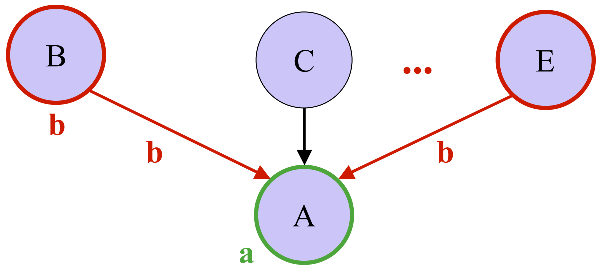
- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

Converging Connections



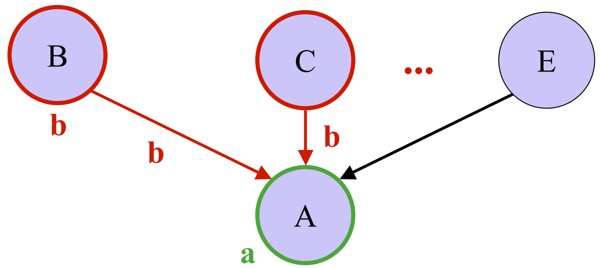
- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

Converging Connections



- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

Converging Connections



- However, if anything is known about the consequence, then information on one possible cause may tell us something about the other causes. (**explaining away**)

Conitional Independence II: d-Separation

Two distinct variables A and B in a Bayesian network are **d-separated** if, for all (undirected) paths between A and B, there is an intermediate variable V (distinct from A and B) such that

- The connection is serial or diverging and V is instantiated, or
- The connection is converging, and neither V nor any of V´s descendants have reveived evidence.

If A and B are not d-separated, then they are called **d-connected**.

d-Separation

If A and B are **d-separated** through the intermediate variable V then

$\mathrm{Ind}(A; B|V)$ (if the connection is serial or diverging and V is instantiated)

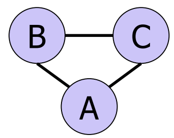
$\mathrm{Ind}(A; B)$ (if the connection is converging, and Neither V nor any of V´s descendants have received evidence)

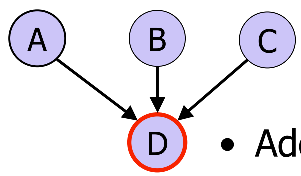
Outline

- Introduction
- Reminder: Probability theory
- Bascis of Bayesian Networks
- Modeling Bayesian networks
- Inference
- Markov Networks
- Learning Bayesian networks
- Relational Models

Undirected Relations

- Dependence relation R among A,B,C
- Neither desirable nor possible to attach directions to them





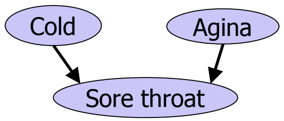
- Add new variable D with states *y*, *n*
- Set

$$P(D = y|A, B, C) = R(A, B, C)$$
$$P(D = n|A, B, C) = 1 - R(A, B, C)$$

- Enter evidence D=*y*

Noisy or

- When a variable A has several parents, you must specify the CPD for each configuration of its parents
  - Number of cases for each configuration in a database to small, configurations to specific for experts, ...



$P(Sore\ throat|Cold)$

$P(Sore\ throat|Agina)$

$P(Sore\ throat|Cold, Agina)$?

Noisy or

- $A_1, \ldots, A_n$ binary variables **listing all the causes** of binary variable B
- Each **event** $A_i = y$ **causes** $B = n$ (**independently** of the other events) **unless an inhibitor prevents it** with probability

$$P(B = n | A_i = y) = q_i$$

- Then

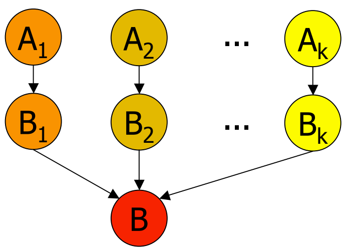
$$P(B = n | A_i, \ldots, A_k) = \prod_{j \in Y} q_i$$

$$P(B = y | A_i, \ldots, A_k) = 1 - \prod_{j \in Y} q_i$$

where Y is the set of indeces of variables in state y

Noisy or

- linear in the number of parents
- $P(B = y | A_i = \ldots = A_k = n) = 0$
- Can be modelled directly:

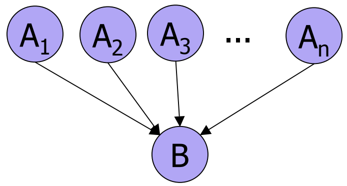


$P(B|A_i)$

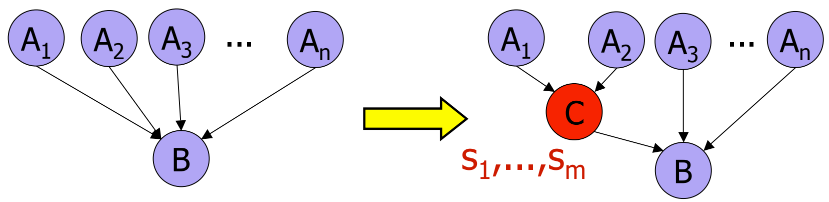
logical or

- Complementary construction is **noisy and**

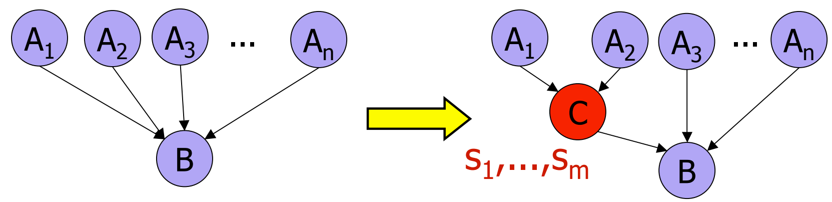
Divorcing



Divorcing



Divorcing



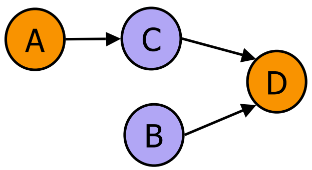
The set of parents $\mathbf{A_1,\ldots,A_i}$ for B are divorced from the parents $\mathbf{A_{i+1},\ldots,A_n}$

Set of configurations $c_1,\ldots,c_k$ of $\mathbf{A_1,\ldots,A_i}$ can be partitioned into the sets $s_1,\ldots,s_m$ s.t. whenever two configurations c,c´ are elements of the same s_i then

$$P(B|c, A_{i+1}, \ldots, A_n) = P(B|c', A_{i+1}, \ldots, A_n)$$

Experts Disagreements

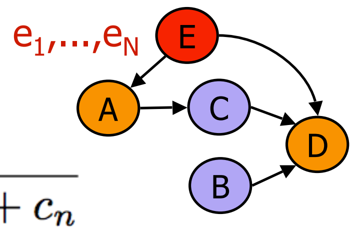
- Expert $\mathbf{e_1,\ldots,e_n}$ disagree on the conditional probabilities of a Bayesian network



- Experts disagree on **A**, **D**
- Confidences into experts $c_1,\ldots,c_n$
  e.g. (2, 1, 5, 3, ..., 1)

Expert disagreements

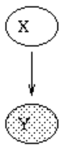
- Introduce a new variable E having the (confidence) distribution $\mathbf{p_1,\ldots,p_n}$ and link it to each variable, the tables of which the experts disagree upon.



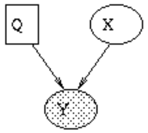
$$p_i = \frac{c_i}{c_1 + c_2 + \cdots + c_n}$$

- The child variables have a CPD for each expert

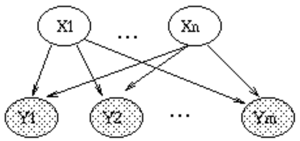
Other Models ...



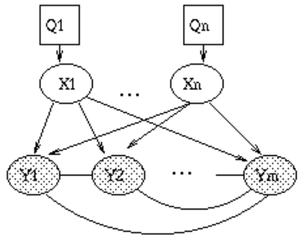
Factor Analysis/PCA



Mixture of FAs



Factor analysis



Independent Factor Analysis