Zurich Research Laboratory

Graphical models and causality

reading group presentation

Isabelle Guyon & Prof. Joachim Buhmann, ETHZ
Andre Elisseeff – IBM Research
7 Mar. 2006

Road Map

- Common sense on causality and uncertainty
- Bayesian Networks
- Causal Networks
- Discussion about co-observations and Causality
- Overview and schedule of the reading group

© 2002 IBM Corporation

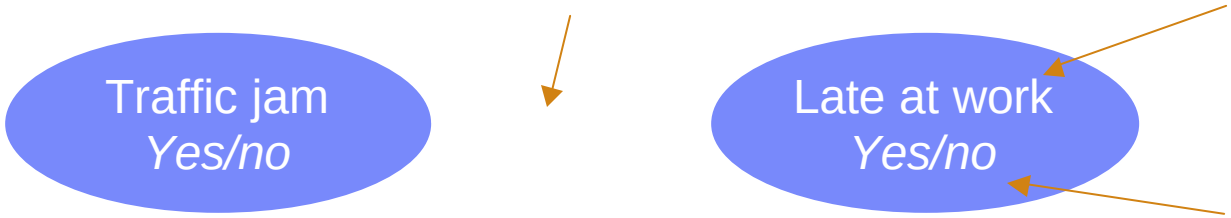
Causality

- Definition: Cause
  "the one, such as a person, an event, or a condition, that is responsible for an action or a result."
- Operational definition:
  X causes Y = changing X will make Y change.

© 2002 IBM Corporation

Common sense causal reasoning

- Causal reasoning can be represented by using the language of graphs:



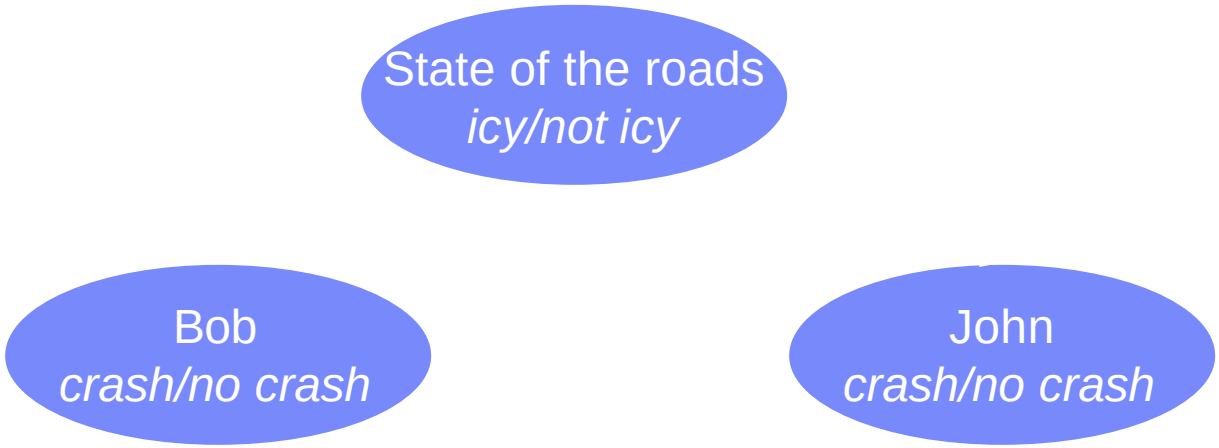
© 2002 IBM Corporation

Icy roads – example of causal reasoning

- Joe is waiting for his two friends Bob and John who are both late for an appointment
- Joe is worried that if the roads are icy one or both of them may have crashed his car.
- Suddenly Joe learns that Bob has crashed
- Joe thinks: If Bob has crashed, probably the roads are icy, then John has probably crashed too!
- Joe then learns it is warm outside and the roads are salted.
- Joe thinks: Bob was unlucky, John will make it.

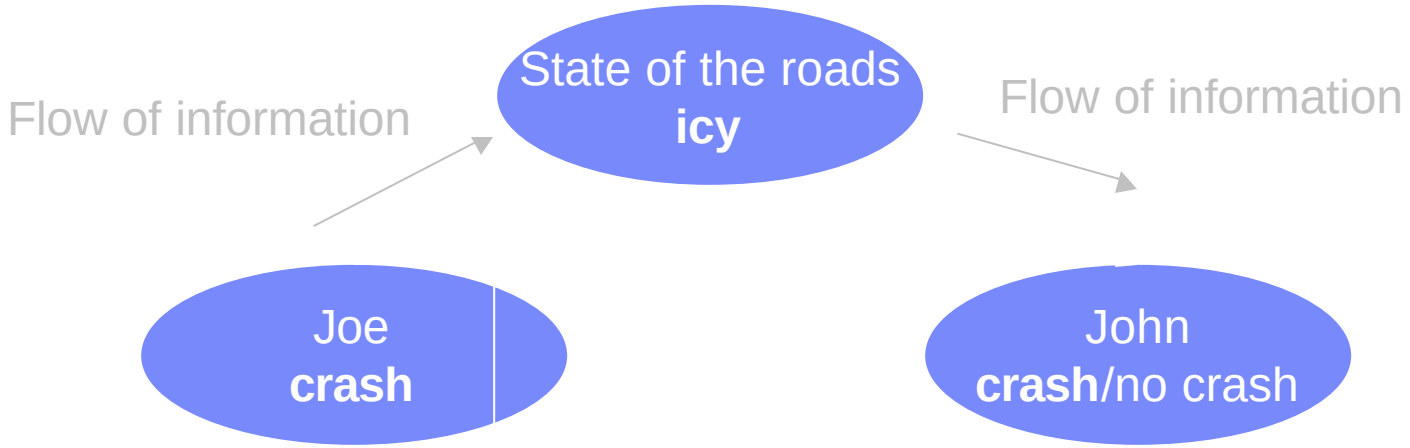
© 2002 IBM Corporation

Simple causal relationships



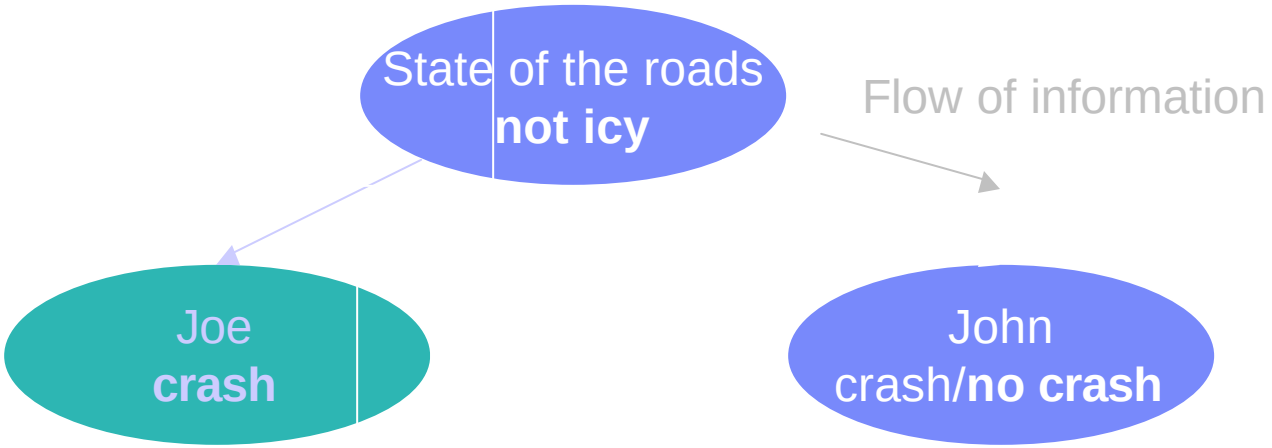
© 2002 IBM Corporation

Watson has crashed!



© 2002 IBM Corporation

But the roads are salted!



© 2002 IBM Corporation

Causality usually goes with probability

- "Causality utterances are often used in situations that are plagued with uncertainty":

"reckless driving causes accident"

"you will fail the course because of your laziness"

1. Causes are likely not certain
2. Causal expressions always have exceptions

© 2002 IBM Corporation

Conditional probabilities – how to express and encode our belief?

- We assume Joe's crash depends on the state of the roads only:

- Without any other information, there is 20% chances that the roads are icy

	icy	not icy
crash	0.2	0.02
no crash	0.8	0.98

Icy	0.2
not icy	0.8

© 2002 IBM Corporation

Conditional probabilities (cont.)

- Joe is worried that if the roads are icy one or both of them may have crashed his car.
) P(crash | icy) = 20% (for both Bob and John)
- Suddenly Joe learns that Bob has crashed
  P(bob crash) = P(bob crash | icy)P(icy) + P(bob crash | not icy) P(not icy)
  = 4.8%
- Joe thinks: If Bob has crashed, probably the roads are icy, then John has probably crashed too!
  P(icy | bob crashed) = P(bob crash | icy)P(icy)/P(bob crash)
  = 0.2*0.2/0.048 = 83%

  P(John crash | bob crash) ¸ P(John crash | icy)P(icy|bob crash)
  ¸ 16.7%
- Joe then learns it is warm outside and the roads are salted.
  P(icy) = 0
- Joe thinks: Bob was unlucky, John will make it.
  P(John crash | bob crash, not icy) = P(John crash | not icy) = 2%

© 2002 IBM Corporation

Conditional probabilities -

- Causal reasoning is based on the calculus of probabilities.
- These probabilities correspond to beliefs.
- They can be updated by data.

) Bayesian interpretation of probabilities

© 2002 IBM Corporation

The relative frequency approach to probability

"A probability is a frequency"

Refers to a "frequentist" approach where random events exist in nature and their statistics can be estimated by counting occurrences.

- Ex.: rolling a dice is a random process.

© 2002 IBM Corporation

The subjective/bayesian approach to probability

- A probability is a degree of belief in my own knowledge of the world.
  Ex.: what are the chances that Ireland wins over France in a soccer game?
- Depends on the subject but can be updated using data.
- Uses Bayes' Theorem to infer unknown probabilities/beliefs from known ones.
- Different interpretation of the world: it is not important to know whether it's random or not. What matters is the uncertainty in our own knowledge.
- Ex.: If you knew everything about the dice and how it was thrown, you could predict the outcome.

© 2002 IBM Corporation

Difference between frequentist and subjective/Bayesian approach

- A probability can refer to:

  Observations in the case of the “frequentist” approach.

  Beliefs (or knowledge) in the case of the “subjective” approach. Beliefs are personal.

- In some cases, a subjective approach is not desired. Ex.: pharmaceutical companies testing the effectiveness of a drug.

© 2002 IBM Corporation

Road Map

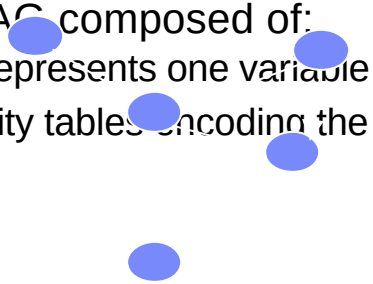
- Common sense on causality and uncertainty
- Bayesian Networks
- Causal Networks
- Discussion about co-observations and Causality
- Overview and schedule of the reading group

© 2002 IBM Corporation

Bayesian network - definition

- A directed acyclic graph (DAG) is a graph without cycles and whose edges are oriented.

- A Bayesian network is a DAG composed of:
  - A set of nodes, each node represents one variable modeling the process of interest.
  - A set of conditional probability tables encoding the local dependencies between the nodes/variables.



© 2002 IBM Corporation

Bayesian network - example

© 2002 IBM Corporation

Bayesian network - properties

- From the graph and the conditional probabilities, one can compute:

  $P(x_1,..,x_n)$ the joint probabilities of all the $x_1,..,x_n$ nodes/variables represented in the network.

  Example:

P(slippery,wet,sprinkler,rain,season) =

P(slippery | wet)P(wet|sprinkler,season)P(sprinkler|season)P(rain|season)P(season)

© 2002 IBM Corporation

Bayesian networks – another definition

- If we know how to compute $P(x_1,..,x_n)$ then we can infer anything we want about the state of the world.
- Problem: “curse of dimensionality”, as n increases, the number of observations needed to compute $P(x_1,..,x_n)$ grows exponentially.
- Cure:

$$P(x_1, .., x_n) = \prod_j P(x_j | x_1, .., x_{j-1})$$

$$P(x_1, .., x_n) = \prod_j P(x_j | \mathrm{pa}_j)$$

© 2002 IBM Corporation

Bayesian networks - properties

- Once a Bayesian network is built, we can infer the joint probability
- and the marginal probabilities

$$P(X_1, .., X_n)$$

- And the conditional probabilities

$$P(X_1, .., X_i) = \sum_{x \notin X_1, .., X_i} P(X_1, .., X_n)$$

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

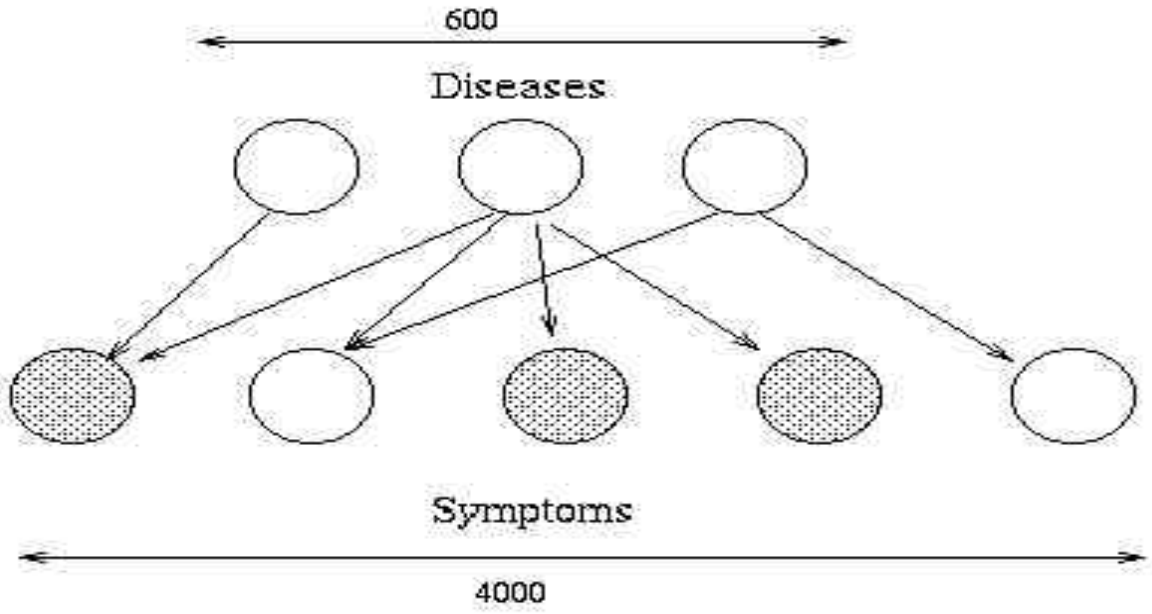
© 2002 IBM Corporation

Bayesian networks (BN)

- BN encodes conditional independence not causality but for most people conditional independence is understood and expressed in causality terms.
- Conditional independence is about observations and co-occurrences.
- Causality is about experimenting.

© 2002 IBM Corporation

Quick Medical Reference (QMR DT)

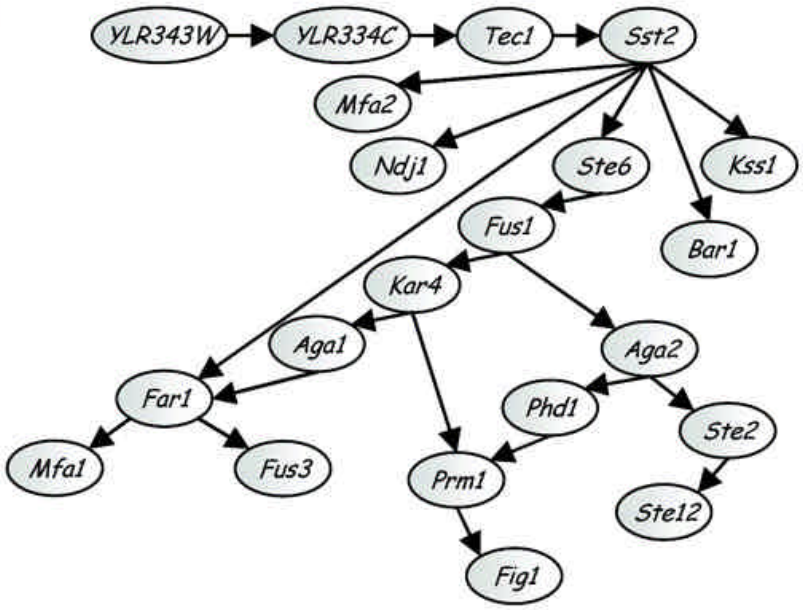


© 2002 IBM Corporation

Gene networks

Infer gene activities from expression data

(*Inferring Cellular Networks Using Probabilistic Graphical Models*, Nir Friedman, Science 2004).



© 2002 IBM Corporation

Road Map

- Common sense on causality and uncertainty
- Bayesian Networks
- Causal Networks
- Discussion about co-observations and Causality
- Overview and schedule of the reading group

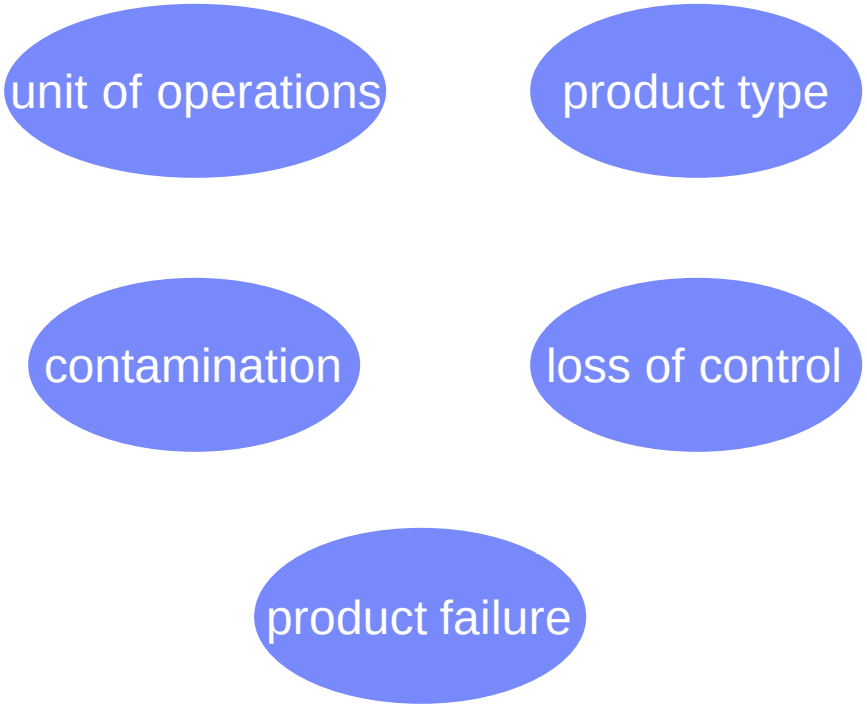
© 2002 IBM Corporation

Causal network - definition

- A causal network is similar to a Bayesian network but:
  - The edge X ! Y exists *iff* X is a direct cause of Y.
  - The conditional probabilities have a different interpretation:
    - P(X|Y) means Y is a cause of X and not only X and Y are related
    - The strength of Y being a cause of X is given by this probability P(X|Y)

© 2002 IBM Corporation

Causal network - example



© 2002 IBM Corporation

Causality networks – Other approaches

- Using Bayesian networks as the underlying probabilistic model for a causal network seems natural.
- Other approaches have been designed where cause and effects are links by functional relationship (Structural Equation Modeling):

$Cause_{N+1} = f(cause_1, cause_2, .., cause_N)$

$effect = g(cause_{N+1}, cause_{N+2}, .., cause_{N+M})$

© 2002 IBM Corporation

Road Map

- Common sense on causality and uncertainty
- Bayesian Networks
- Causal Networks
- Discussion about co-observations and Causality
- Overview and schedule of the reading group

© 2002 IBM Corporation

The Simpson's paradox

- Consider the following study on the effect of a drug C:
  - Overall recovery rate is higher when the drug is taken

  - When separated into two groups (male/female):
    - Recovery rate for male is higher when the drug is not taken
    - Recovery rate for female is also higher when the drug is not taken.

© 2002 IBM Corporation

Conditional probabilities

$$P(R|C) > P(R|\neg C)$$

$$P(R|C,F) < P(R|\neg C,F)$$
$$P(R|C,\neg F) < P(R|\neg C,\neg F)$$

© 2002 IBM Corporation

Example

	Recovered	Not recovered		Recovery rate
Drug (C)	20	20	40	50%
No drug (: C)	16	24	40	40%
	36	44	80	

	Recovered	Not recovered		Recovery rate
Drug (C)	18	12	30	60%
No drug (: C)	7	3	10	70%
	36	44	40	

	Recovered	Not recovered		Recovery rate
Drug (C)	2	8	10	20%
No drug (: C)	9	21	30	30%
	11	29	40	

© 2002 IBM Corporation

Existence of a confounding factor

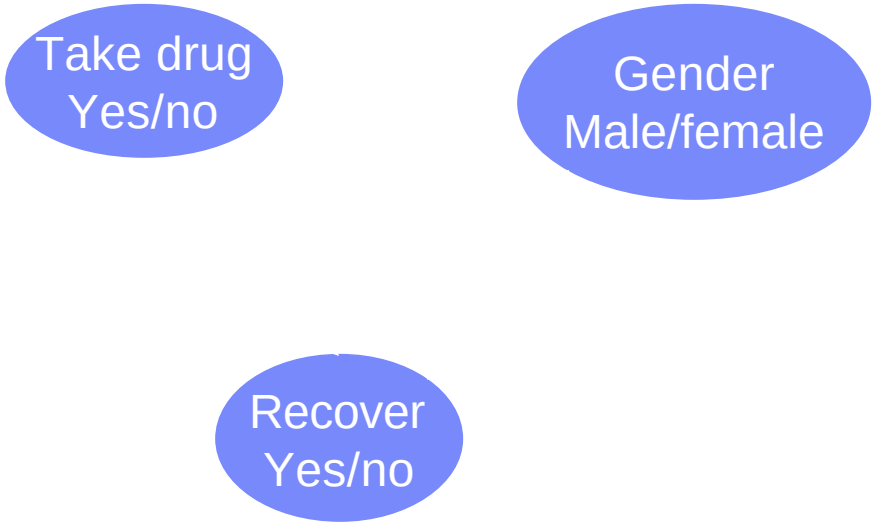
- Male patients recover more often (regardless of the drug) and are more likely than the females to use the drug
) bias the statistics toward positive results unless the population is split.

Correct conclusion: not taking the drug is better.

© 2002 IBM Corporation

What is behind this confounding factor?

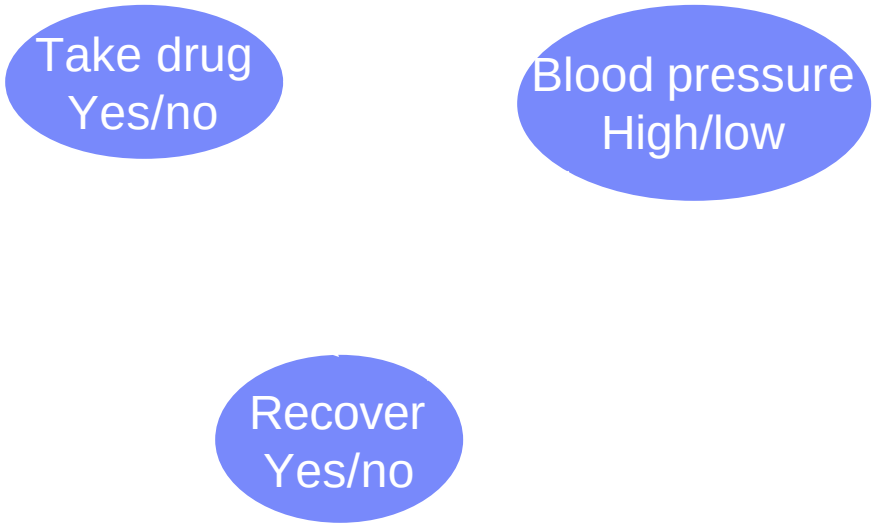
- We interpret the data in terms of the following causal diagram



© 2002 IBM Corporation

Other interpretation

- Change gender by blood pressure, then the recovery rate for the whole population makes sense.



© 2002 IBM Corporation

Implicit causal reasoning when we speak about conditional probabilities

- There is no statistical way to remove confounding factors (external factors that make appear spurious associations)
- One way is the causal way: need to import/consider causality when building a statistical model.

© 2002 IBM Corporation

Correlation is not causation

- “The increase of Dow Jones is correlated with the length of my hair”: they both increase regularly. It does not mean they are related.
- Question: how often does an engineer perform root cause analysis with correlation? () implicit causal network?)

© 2002 IBM Corporation

Road Map

- Common sense on causality and uncertainty
- Bayesian Networks
- Causal Networks
- Discussion about co-observations and Causality
- Overview and schedule of the reading group

© 2002 IBM Corporation

Goal of this reading group

- Understand the key concepts of causality and their links with statistics.
- Main thread will be the reading of 5 chapters from the book of Judea Pearl: “Causality, Reasoning and Inference”.
- Practical understanding will be supported by the use of the Genie software.

© 2002 IBM Corporation

Schedule 1 -

	Dates (Tuesday)	Theme	Paper	Discussion leader	Summary, discussion, and/or slides
1	4 April	Introduction	Introduction to probabilistic reasoning (tutorial)	A. Elisseeff	
2	11 April	I. Basic concepts	Bayesian models without tears, Eugene Charniak	I. Guyon	Install Genie. Build an example network with the software implementing the example of Fig. 2 of the paper.
3	18 April		**Probabilistic Reasoning.** Chapter 14 of the book "Artificial Intelligence: A modern approach" by Stuart Russell and Peter Norvig.		Slides a. Slides b.
4	25 April		**Introduction to probabilities, graphs, and causal models**. Chapter 1 of the book "Causality" by Judea Pearl. The whole chapter is available in pdf.		Slides

© 2002 IBM Corporation

Schedule 2 -

5	2 May	II. Basic methods	**Belief propagation.** An introduction to factor graphs, by Hans-Andrea Loeliger.		
6	9 May		**Structure learning.** A tutorial on learning with Bayesian networks, by David Heckerman.		
7	16 May		**Variational methods.** An Introduction to Variational Methods for Graphical Models, by Michael Jordan, Zoubin Ghahramani, Tommi Jaakkola, and Lawrence Saul.		
8	23 May		**A theory of inferred causation.** Chapter 2 of the book "Causality" by Judea Pearl. Introduction in pdf.		Slides

© 2002 IBM Corporation