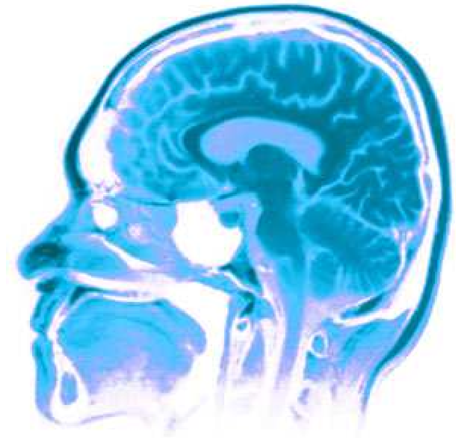
CPSC340



Probabilistic Graphical Models

Nando de Freitas
September, 2012
University of British Columbia

Outline of the lecture

Thi lecture introduces probabilistic graphical models. These are extremely powerful representations that enable us to scale probabilistic models to large real domains. The objective is to learn the following topics:

- ❑ The curse of dimensionality
- ❑ Definition of a DAG (aka Bayesian network)
- ❑ Conditional independence in DAGs
- ❑ Domains of application of Bayesian nets.

The curse of dimensionality

This curse tells us that to represent a joint distribution of **d** *binary variables, we need* $(\mathbf{2^d})-1$ *terms!*

$d=3$

A	B	C	
0	0	0	P(A=0, B=0, C=0)
0	0	1	P(A=0, B=0, C=1)
0	1	0	
0	1	1	⋮
1	0	0	
⋮			

7

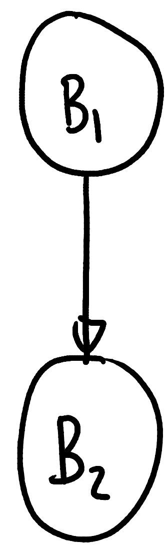
A simple probabilistic graphical model

Let us revisit our problem of drawing balls from the set **{r,r,r,b}.**

$$P(B_1) = \begin{array}{c|c} \multicolumn{2}{c}{B_1} \\ r & b \\ \hline 3/4 & 1/4 \end{array}$$

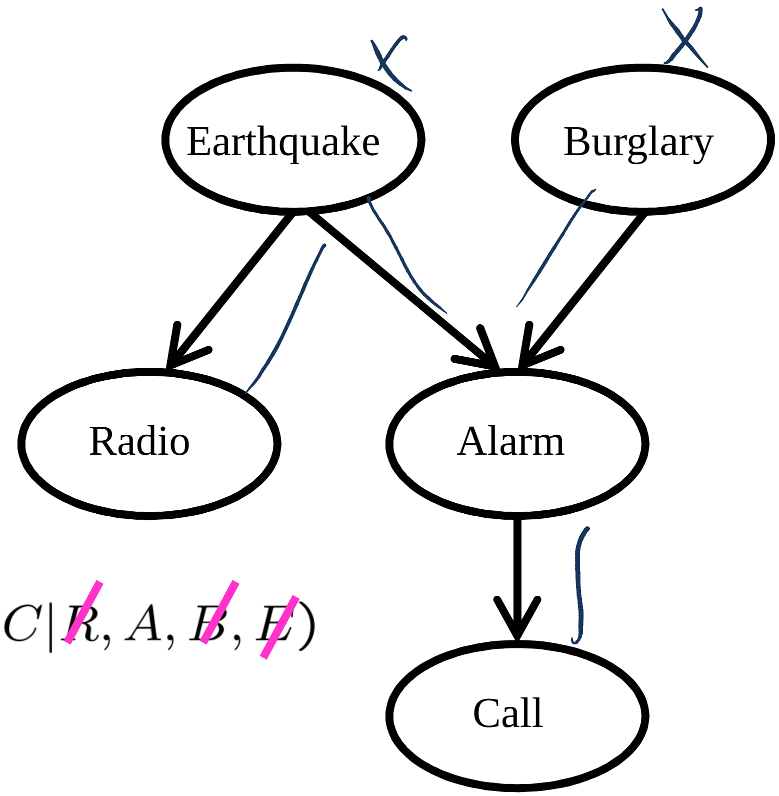
$$P(B_2 \mid B_1) = \begin{array}{cc|c|c} & & \multicolumn{2}{c}{B_2} \\ & & r & b \\ \hline B_1 & r & 2/3 & 1/3 \\ & b & 1 & 0 \end{array}$$

$$P(B_2, B_1) = P(B_1, B_2) = \begin{array}{cc|c|c} & & \multicolumn{2}{c}{B_2} \\ & & r & b \\ \hline B_1 & r & 1/2 & 1/4 \\ & b & 1/4 & 0 \end{array}$$



Directed probabilistic graphical models

1. Directed Acyclic Graph (DAG)
   Nodes – random variables
   Edges – direct influence ("causation")
2. X_i independent of $X_{\text{ancestors}} \mid X_{\text{parents}}$



drop

$$
\begin{aligned}
& P(B, E, A, R, C) \\
= \; & P(B)P(E|\cancel{B})P(A|B,E)P(R|\cancel{A},\cancel{B},E)P(C|\cancel{R},A,\cancel{B},\cancel{E}) \\
= \; & P(B)P(E)P(A|B,E)P(R|E)P(C|A)
\end{aligned}
$$

The DAG tells us that if we have ***n*** *variables* x_i*, the joint distribution of these variables* ***factorizes*** *as follows:*

$$\boldsymbol{P(x_{1:n}) = P(x_1, x_2, \ldots, x_n) =} \prod_{i=1}^{n} P(x_i \mid \text{parents}(x_i))$$

Joint vs Factorized joint distributions

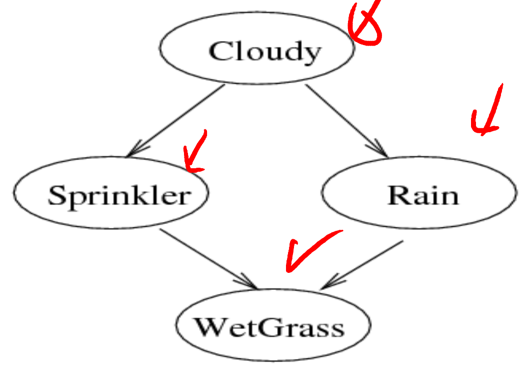
c	s	r	w	prob
0	0	0	0	0.200
0	0	0	1	0.000
0	0	1	0	0.005
0	0	1	1	0.045
0	1	0	0	0.020
0	1	0	1	0.180
0	1	1	0	0.001
0	1	1	1	0.050
1	0	0	0	0.090
1	0	0	1	0.000
1	0	1	0	0.036
1	0	1	1	0.324
1	1	0	0	0.001
1	1	0	1	0.009
1	1	1	0	0.000
1	1	1	1	0.040

16-1=15

$$p(C,S,R,W) = \underline{p(C)}\,\underline{p(S|C)}\,\underline{p(R|C)}\,\underline{p(W|S,R)}$$

P(C=F)	P(C=T)
0.5	0.5

P(c)



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

P(s|c)

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

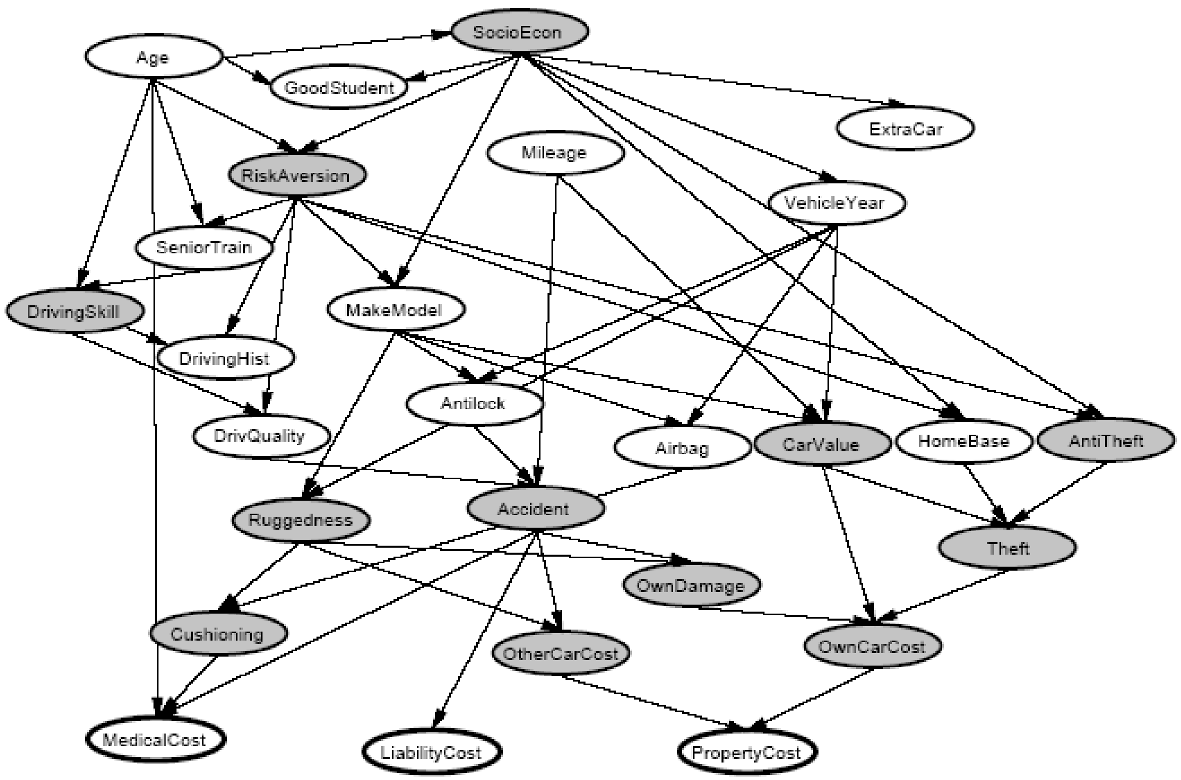
P(R|c)

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

P(w|R,s)

4+2+2+1 = 9

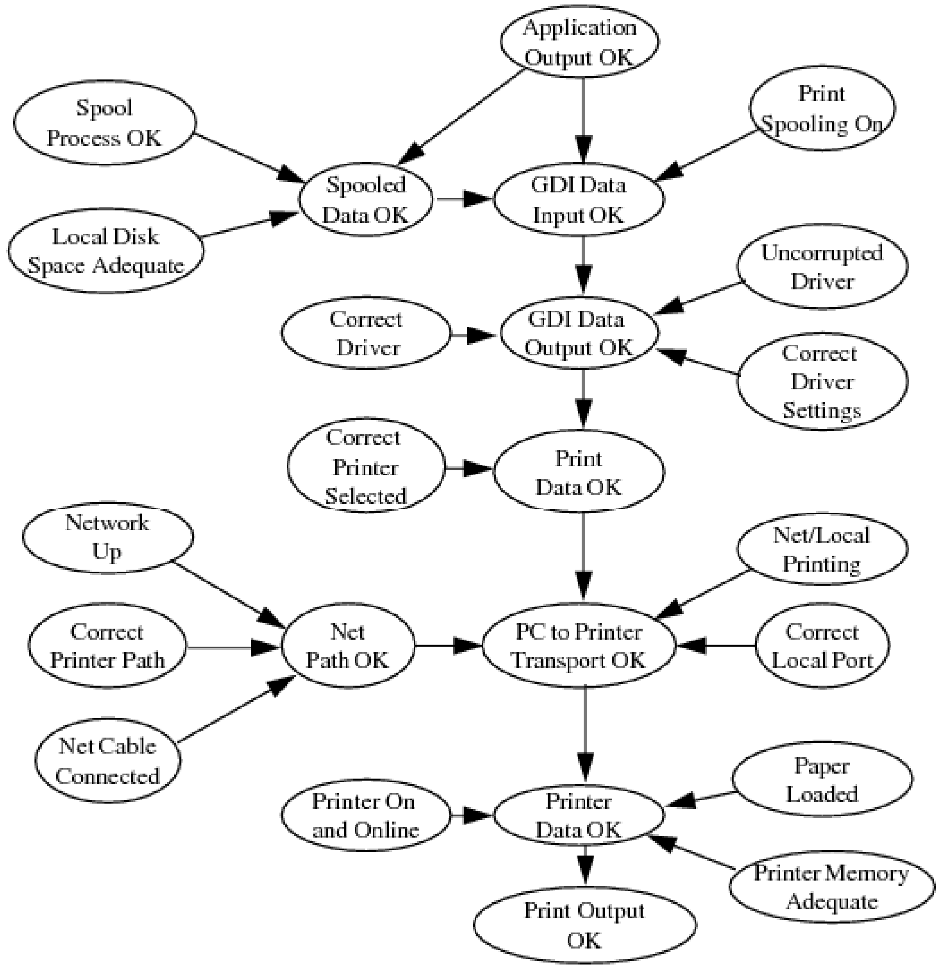
Example: Vehicle insurance



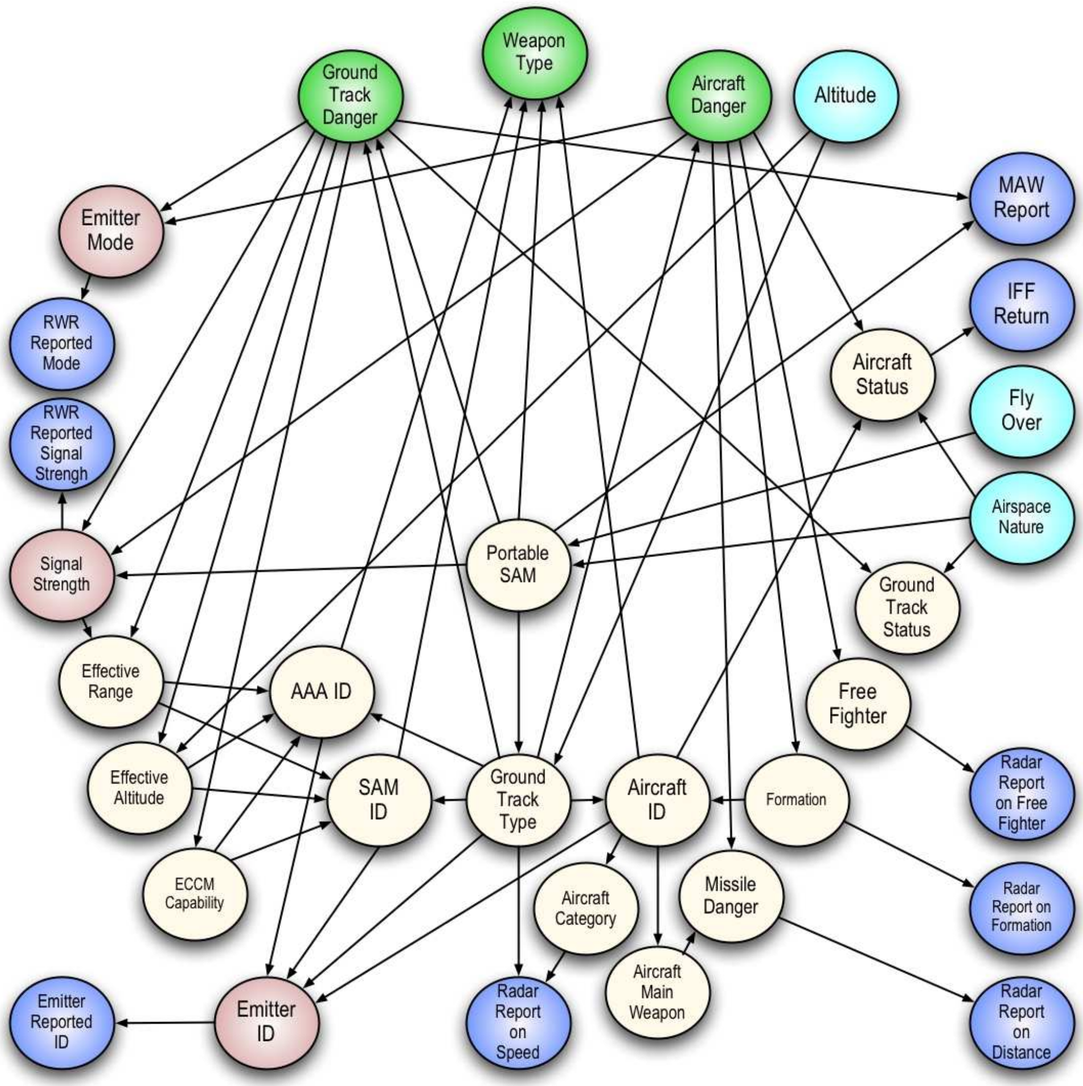
The 12 **shaded variables** are considered **hidden** or **unobservable**, while the other 15 are observable. The network has over 1400 parameters. An insurance company would be interested in predicting the bottom three "cost" variables, given all or a subset of the other observable variables.

http://www.cs.princeton.edu/courses/archive/fall10/cos402/assignments/bayes/

Example: Microsoft Windows Printer Troubleshooter

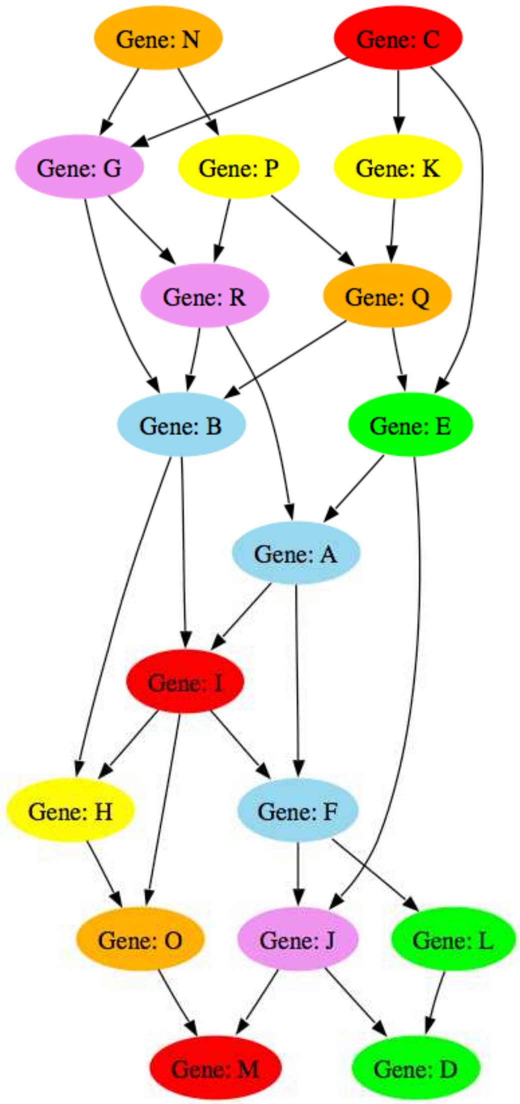


Example: Radar and aircraft control



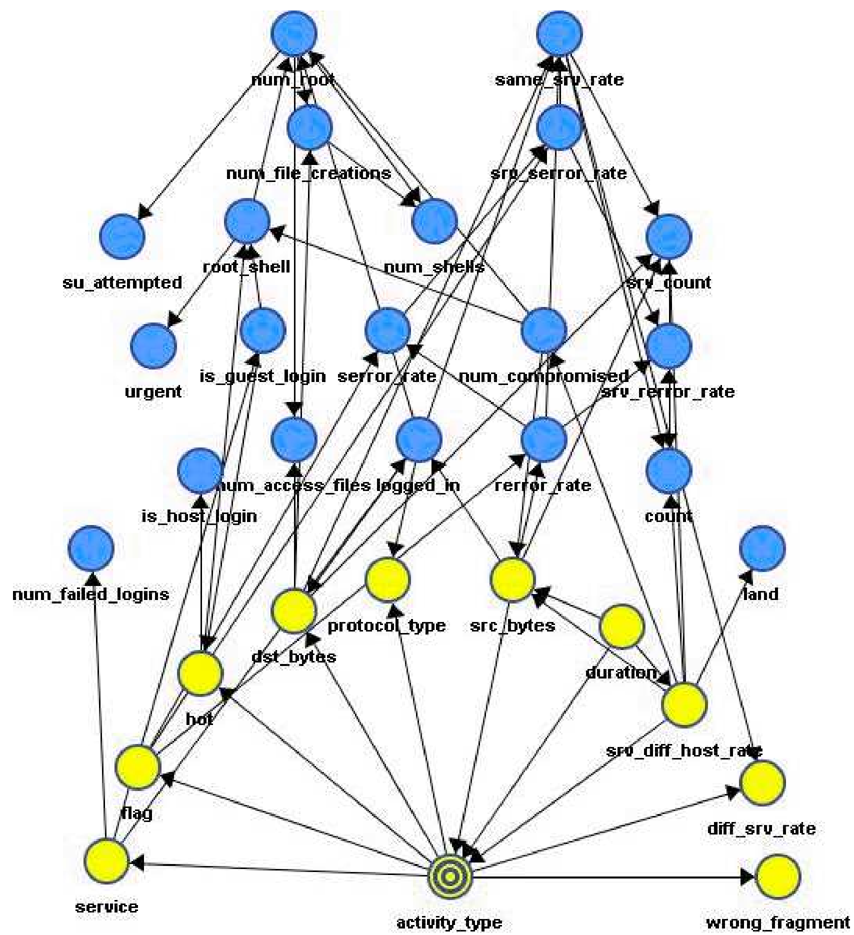
http://www.pr-owl.org/basics/bn.php

Example: Gene expression



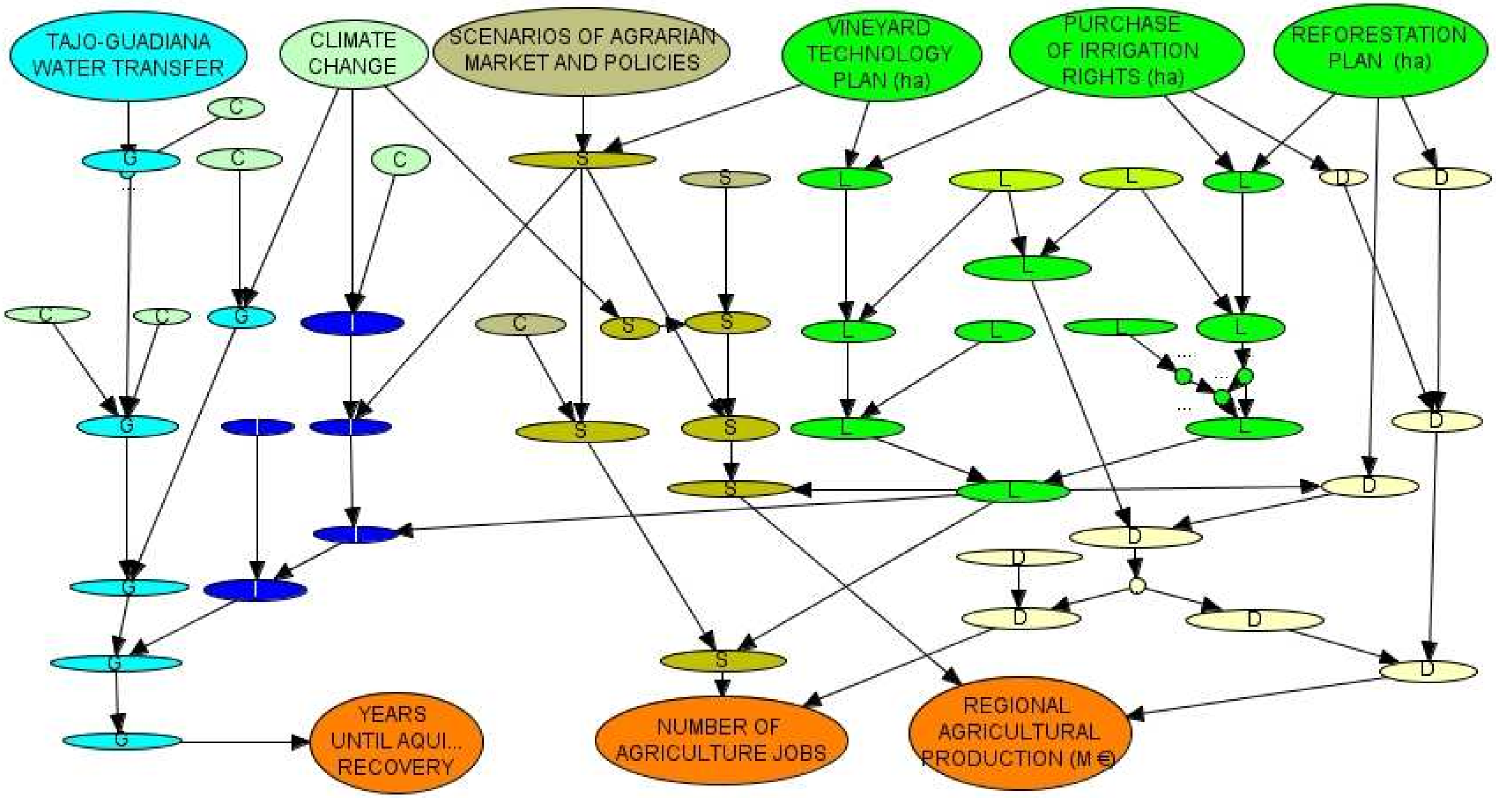
http://www.mcdb.lsa.umich.edu/labs/schiefel/rootepinet/software.php

Example: Cyber crimes detection



http://www.bayesia.com/en/products/bayesialab/resources/case-studies/cyber-crime-detection.php

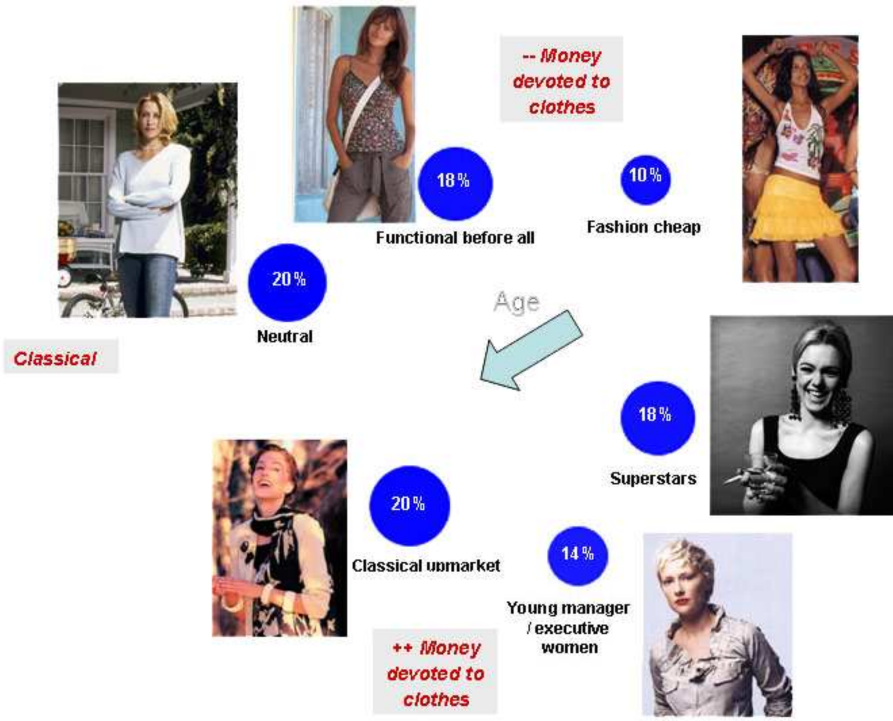
Example: Water resource management

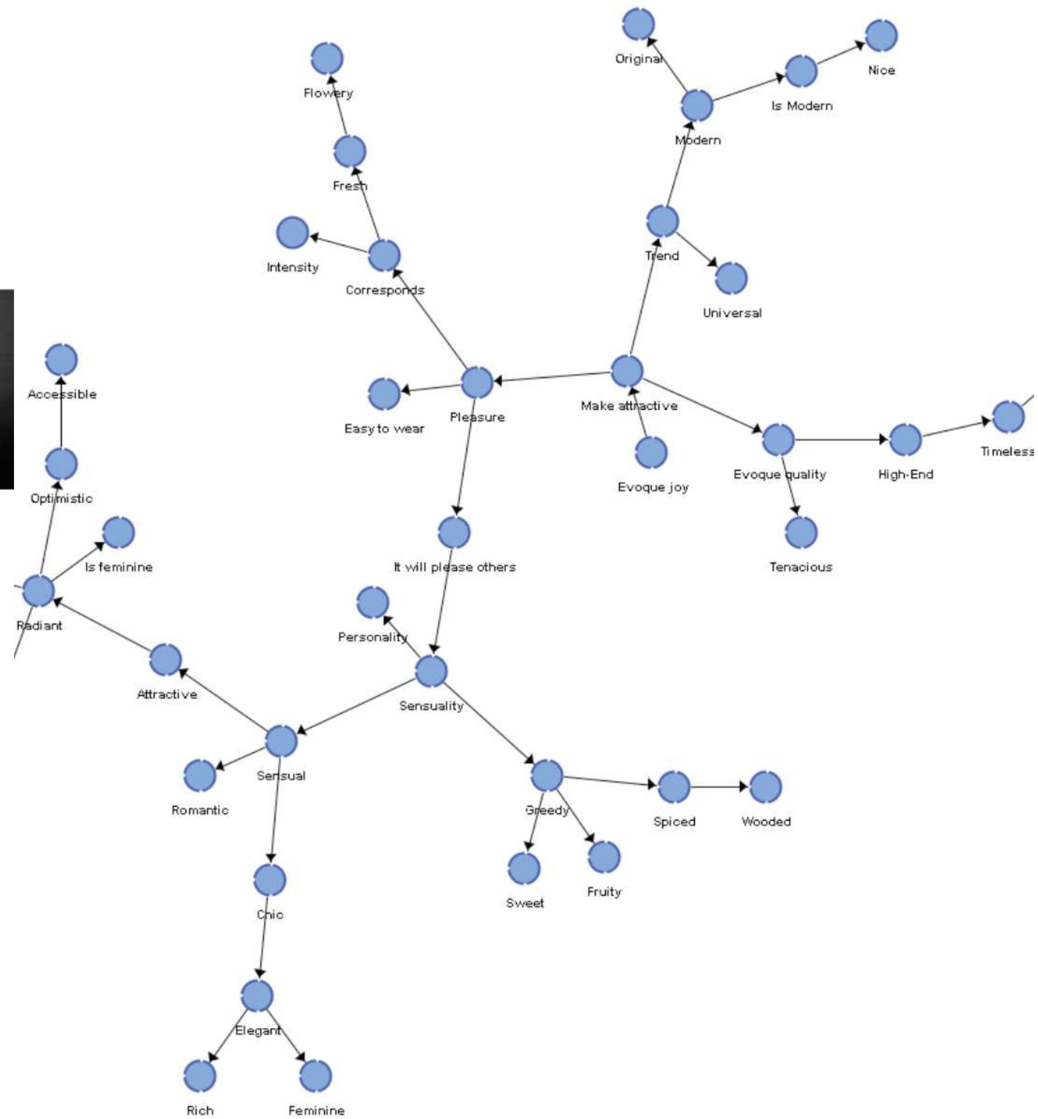


Bayesian network of the Upper Guadiana basin at aquifer-scale (Zorrilla 2009).

"G and light blue" variables refer to **groundwater**; "I and blue" represent **irrigation variables**; "C and light green" variables correspond to **climate**; "S and brown" variables represent **socio-economic scenarios**; "L and green" variables relate to **irrigated land**; and "D and yellow" variables represent **rain-fed agriculture**.

Example: Marketing & fashion

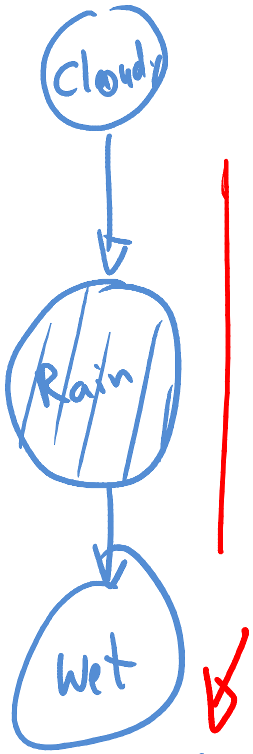




http://www.bayesia.com/en/applications/marketing.php

3 cases of conditional independence to remember

$P(c)P(R|c)P(W|R)$

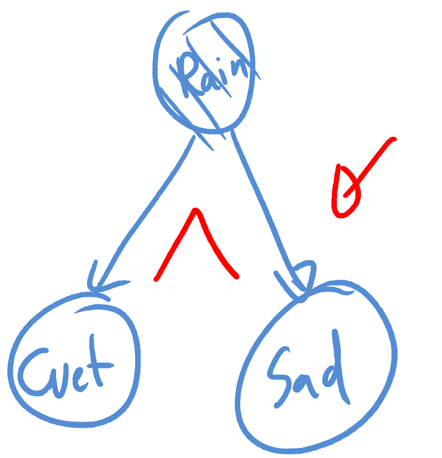


$(W \perp C) \mid R$

given / assuming

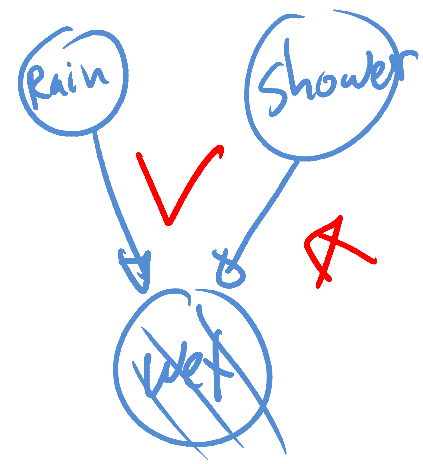
independent

$P(R)P(W|R)P(S|R)$



$(W \perp S) \mid R$

$P(r)P(s)P(w|r,s)$



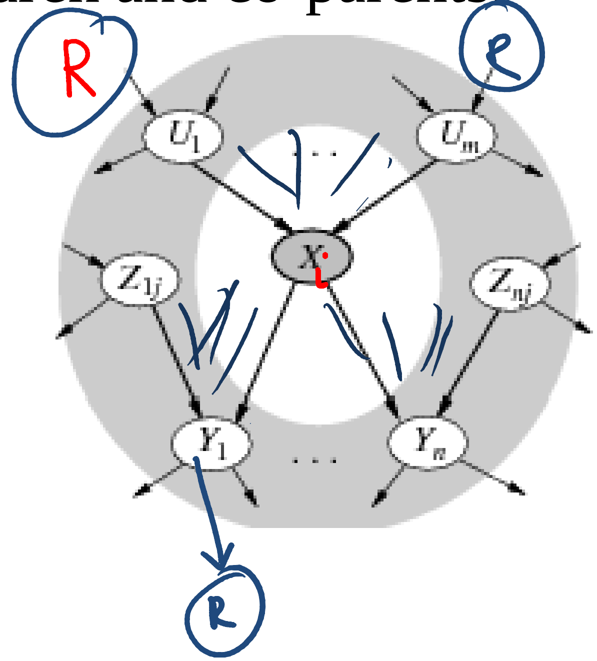
$(R \not\perp S) \mid W$

not!

Markov blankets

$P(X_{-i}) = \sum_{x_i} P(x_i, X_{-i})$

The **Markov blanket** of a node is the set that renders it independent of the rest of the graph. This is the parents, children and co-parents.



node x_i

Claim: $P(x_i | X_{-i}) = P(x_i | U, Y, Z)$

X_{-i} = {all other nodes} = $\{R, U, Y, Z\} = X_{-i}$

Proof sketch

$$P(x_i | X_{-i}) = \frac{P(x_i, X_{-i})}{P(X_{-i})} = \frac{P(x_i, R, U, Y, Z)}{\sum_{x_i=0}^{1} P(x_i, R, U, Y, Z)}$$

$$= \frac{P(x_i|U)\left\{\prod_j P(y_j | x_i, Z)\right\} \cancel{P(Z, R, U)}}{\sum_{x_i=0}^{1} P(x_i|U)\left\{\prod_j P(y_j | x_i, Z)\right\} \cancel{P(Z, R, U)}}$$

$$= P(x_i|U) \prod_j P(y_j | x_i, Z) \Big/ \sum_{x_i} P(x_i|U) \prod_j P(y_j | x_i, Z)$$

Next lecture

In the next lecture, we expand on the topic of inference.