Bayesian Techniques in Vision and Perception

Dr. Olivier Aycard
E-Motion Research Group
GRAVIR-IMAG & INRIA RA
Grenoble, FRANCE

http://emotion.inrialpes.fr/aycard

Dr. Luis Enrique Sucar
Computer Science Dept.
INAOE
Puebla, MEXICO

http://ccc.inaoep.mx/~esucar/

Olivier.Aycard@imag.fr
esucar@inaoep.mx

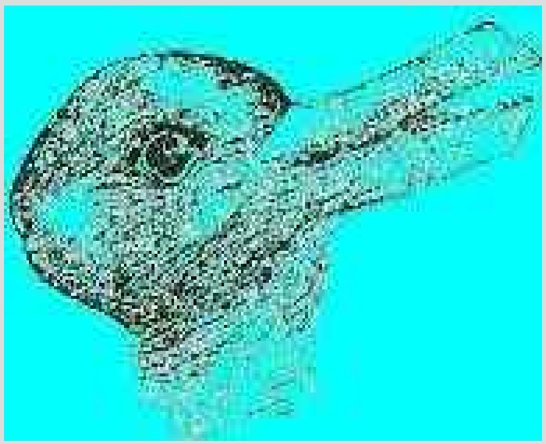
Content

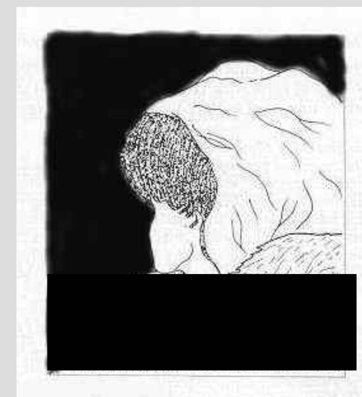
- Fundamentals of Bayesian Techniques (E. Sucar)
- Bayesian Filters (O. Aycard)
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

Content

- Fundamentals of Bayesian Techniques (E. Sucar)
  - Introduction
  - Fundamentals
  - Bayesian Classifiers
  - Bayesian Networks
- Bayesian Filters (O. Aycard)
- Research activities in Vision (E. Sucar)
- Research activities in Perception (O. Aycard)

What do you see?





What we see depends on our previous knowledge (model) of the world

Olivier.Aycard@imag.fr
esucar@inaoep.mx

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

UNIVERSITE JOSEPH FOURIER

Contents

- Fundamentals of Bayesian Techniques
  - Introduction
  - Fundamentals
  - Bayesian Classifiers
  - Bayesian Networks

Bayesian visual perception

- The perception problem is characterized by two main aspects:
  - The properties of the world that is observed (prior knowledge)
  - The image data used by the observer (data)
- The Bayesian approach combines these two aspects which are characterized as probability distributions

Representation

- Scene properties – *S*
- Model of the world – prior probability distribution – $P(S)$
- Model of the image – probability distribution of the image given de scene (likelihood) – $P(I|S)$

Recognition

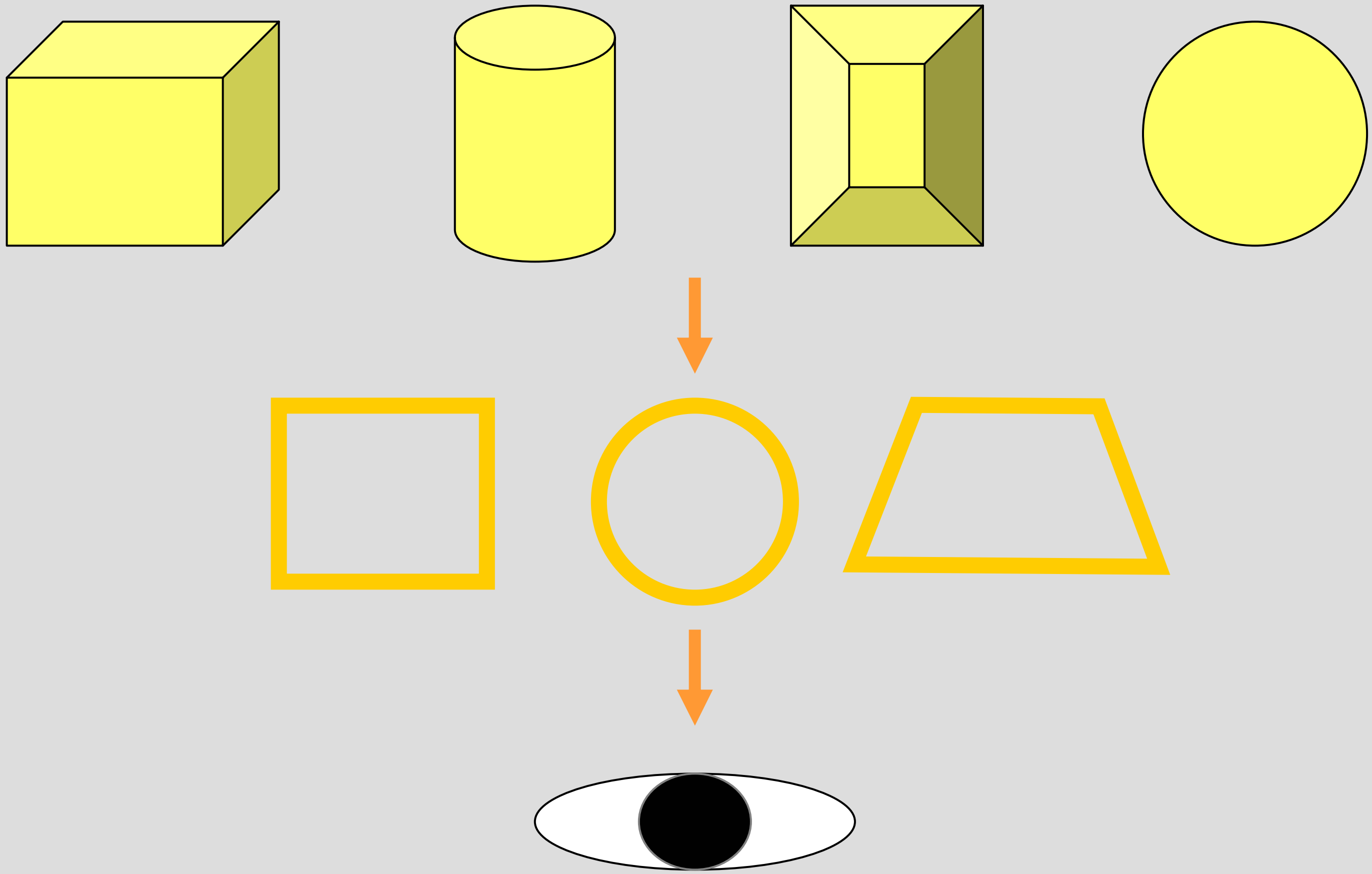
- The scene (object) is characterized by the posterior probability distribution – $P(S|I)$
- By Bayes theorem:

$$P(S|I) = P(S)\, P(I|S) \,/\, P(I)$$

- The denominator can be consider as a normalizing constant:

$$P(S|I) = k\, P(S)\, P(I|S)$$

Example



Example

- Prior distribution of objects – P(O)
  - Cube 0.2
  - Cylinder 0.3
  - Sphere 0.1
  - Prism 0.4

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Example

- Likelihood function P(Silhouette|Object) – P(S|O)

	Cube	Cylinder	Sphere	Prism
Square	1.0	0.6	0.0	0.4
Circle	0.0	0.4	1.0	0.0
Trapezoid	0.0	0.0	0.0	0.6

Example

- Posterior distribution P(Object|Silhouette) – P(O|S)
- Bayes rule:

  $$P(O|S) = k\, P(O)\, P(S|O)$$

- For example, given S=square

  P(Cube | square)= k 0.2 * 1 = k 0.2 = 0.37

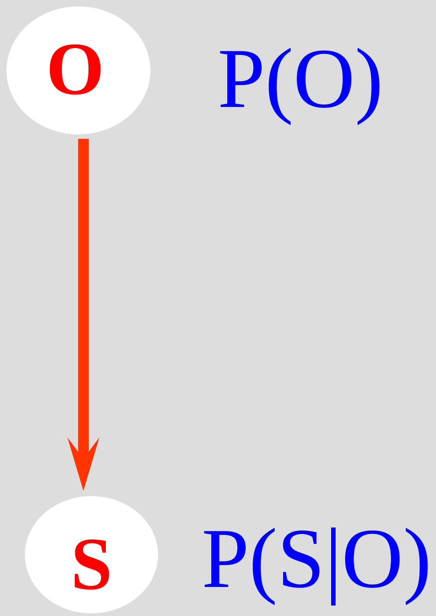
  P(Cylinder | square)= k 0.3 * 0.6 = k 0.18 = 0.33

  P(Sphere | square)= k 0.1 * 0 = 0

  P(Prism | square)= k 0.4 * 0.4 = k 0.16 = 0.30

Graphical Model

- We can represent the dependence relation in this simple example graphically, with 2 variables and an arc



Graphical Models

- This graphical representation of probabilistic models can be extended to more complex ones.
- There are several types of probabilistic graphical models (PGMs) that can be applied to different problems in vision
- We first review PGMs and then introduce some models and their application in vision

Contents

- Fundamentals of Bayesian Techniques
  - Introduction
  - Fundamentals
  - Bayesian Classifiers
  - Bayesian Networks

Probabilistic Graphical Models

- Given a set of (discrete) random variables,

$$\boldsymbol{X} = X_1, X_2, \ldots, X_N$$

- The joint probability distribution,

$$P(X_1, X_2, \ldots, X_N)$$

- specifies the probability for each combination of values (the joint space). From it, we can obtain the probability of a variable(s) (marginal), and of a variable(s) given the other variables (conditional)

Probabilistic Graphical Models

- A Probabilistic Graphical Model is a compact representation of a joint probability distribution, from which we can obtain marginal and conditional probabilities
- It has several advantages over a "flat" representation:
  - It is generally much more compact (space)
  - It is generally much more efficient (time)
  - It is easier to understand and communicate
  - It is easier to build (from experts) or learn (from data)

Probabilistic Graphical Models

- A graphical model is specified by two aspects:
  - A Graph, *G(V,E)*, that defines the structure of the model
  - A set of local functions, $f(Y_i)$, that defines the parameters (probabilities), where Y_i is a subset of X
- The joint probability is defined by the product of the local functions:

$$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{n} f(Y_i)$$

Probabilistic Graphical Models

- This representation in terms of a graph and a set of local functions (called potentials) is the basis for *inference* and *learning* in PGMs
  - **Inference:** obtain the marginal or conditional probabilities of any subset of variables ***Z*** given any other subset ***Y***
  - **Learning:** given a set of data values for ***X*** (that can be incomplete) estimate the structure (graph) and parameters (local function) of the model

Probabilistic Graphical Models

- We can classify graphical models according to 3 dimensions:
  - Directed vs. Undirected
  - Static vs. Dynamic
  - Generative vs. Conditional

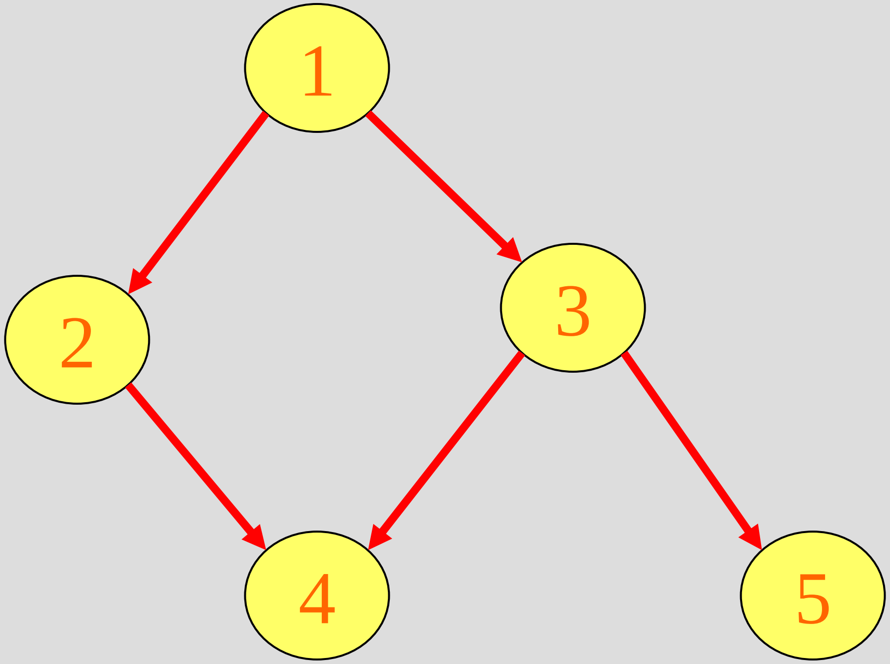
Olivier.Aycard@imag.fr
esucar@inaoep.mx

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

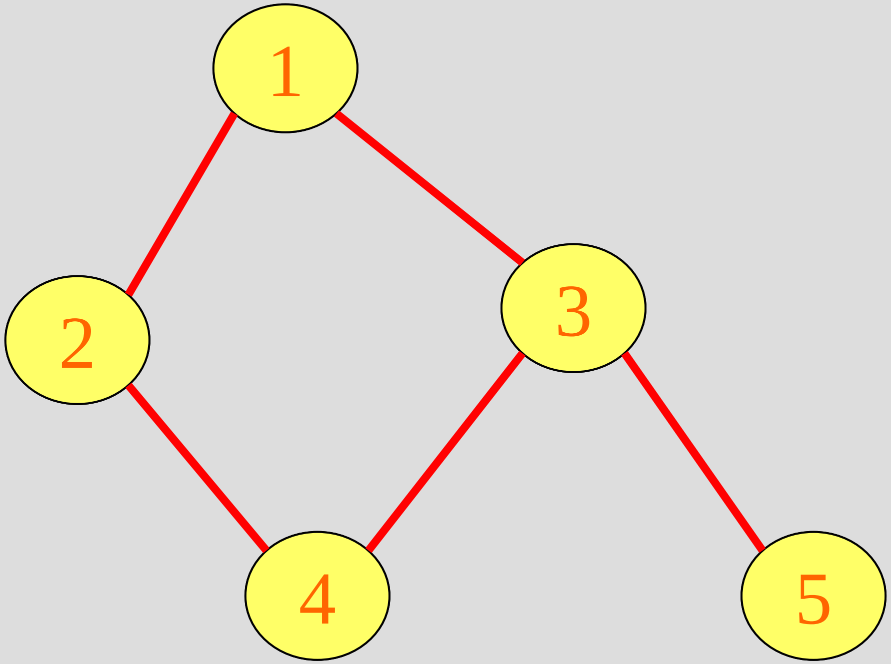
UNIVERSITE JOSEPH FOURIER

Probabilistic Graphical Models

- Directed

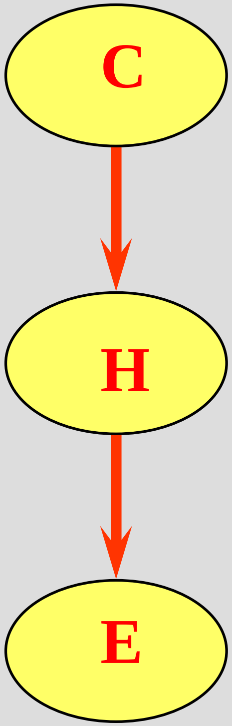


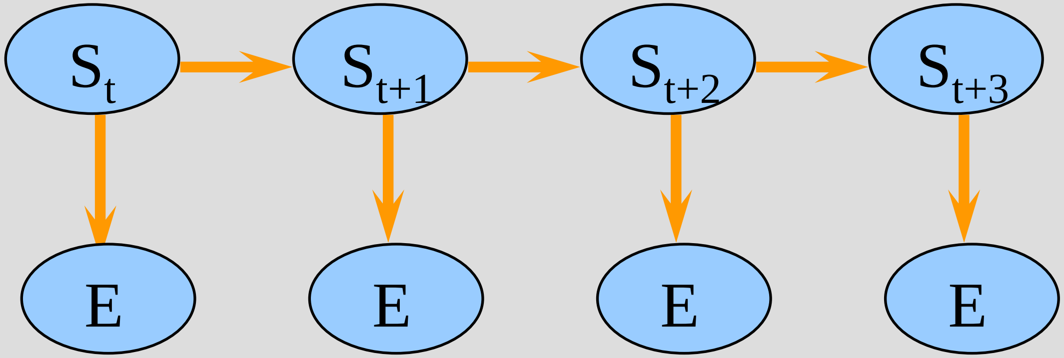
- Undirected



Probabilistic Graphical Models

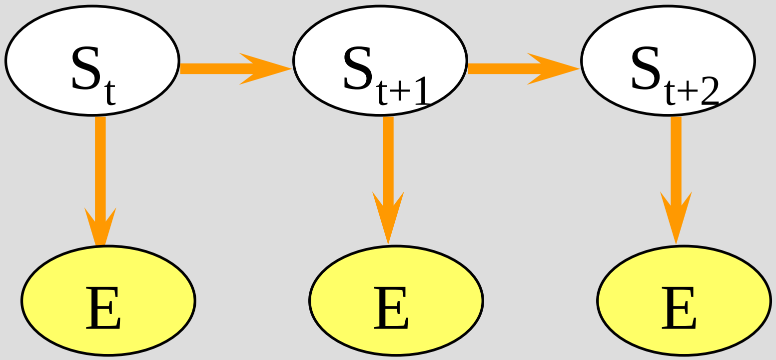
- Static
- Dynamic



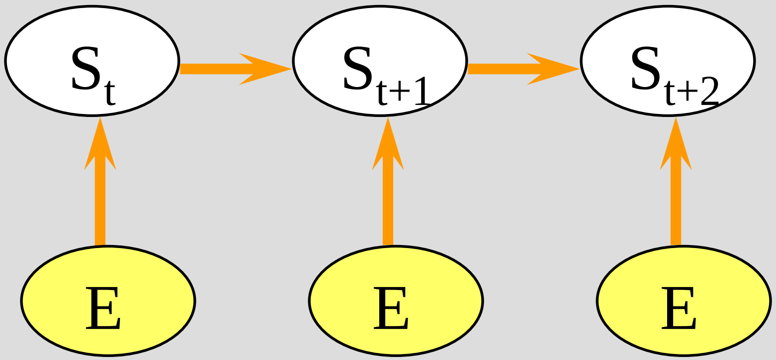


Probabilistic Graphical Models

- Generative



- Conditional



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Types of PGMs

- We will consider the following models and their applications in vision and robotics:
  - Bayesian classifiers
  - Bayesian networks
  - Hidden Markov models
  - Dynamic Bayesian networks
  - Kalman filters
  - Particle filters

Contents

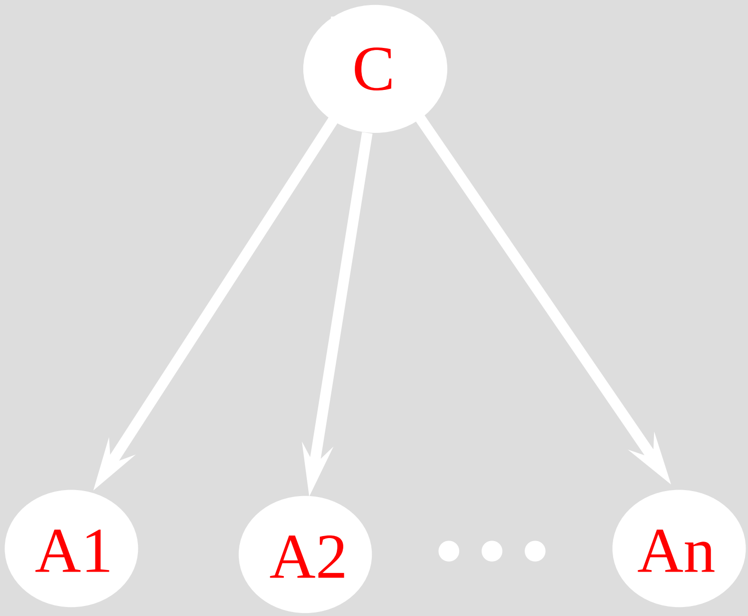
- Fundamentals of Bayesian Techniques
  - Introduction
  - Fundamentals
  - Bayesian Classifiers
  - Bayesian Networks

Bayesian Classifier

- A Bayesian classifier is used to obtain the probability of certain variable (the class or hypothesis, H) given a set of variables known as the attributes or evidence ($\mathbf{E} = E_1, \ldots, E_N$)
- It is usually assumed that the attributes are independent given the class – **Naive Bayesian Classifier** – so its PGM is represented as a "star" with the class as the root and the attributes as the leafs

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Naive Bayesian Classifier



Bayesian Classifier

- The posterior probability of each hypothesis (H) based on the Evidence (E) is:

$$P(H \mid \mathbf{E}) = P(H)\, P(\mathbf{E} \mid H) / P(\mathbf{E})$$

- Usually the exact value of $P(H|E)$ is not required, just the most probable value of H

Naive Bayesian classifier
Inference

- Consider each attribute independent given the hypothesis:

$$P(E_1, E_2, \ldots E_N \mid H) = P(E_1 \mid H)\, P(E_2 \mid H) \ldots P(E_N \mid H)$$

- So the posterior probability is given by:

$$P(H \mid E_1, E_2, \ldots E_N) = [P(H)\, P(E_1 \mid H)\, P(E_2 \mid H) \ldots P(E_N \mid H)] / P(\mathbf{E})$$

$$= k\, P(H)\, P(E_1 \mid H)\, P(E_2 \mid H) \ldots P(E_N \mid H)$$

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Naive Bayesian classifier
Learning

- Structure:
  - the structure is given by the naive Bayes assumption
- Parameters:
  - we need to estimate the prior probability of each class

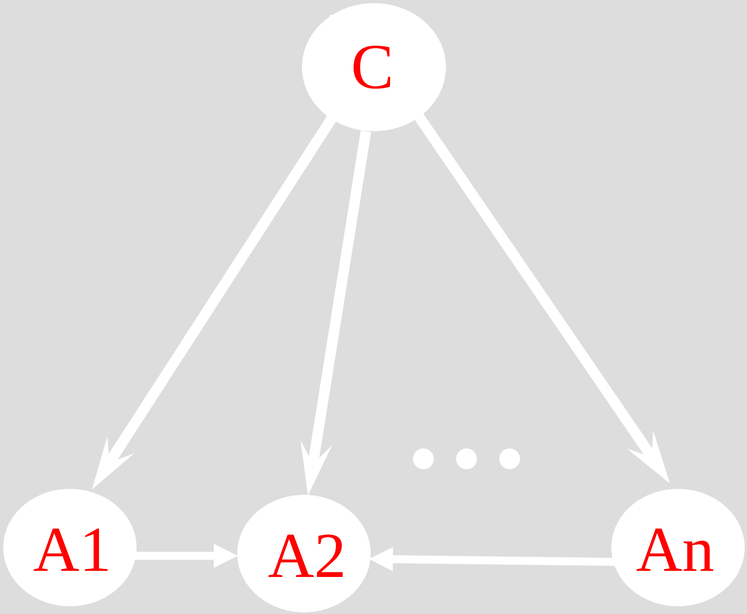
$$P(C_i)$$

  - and the individual conditional probabilities of each attribute given the class

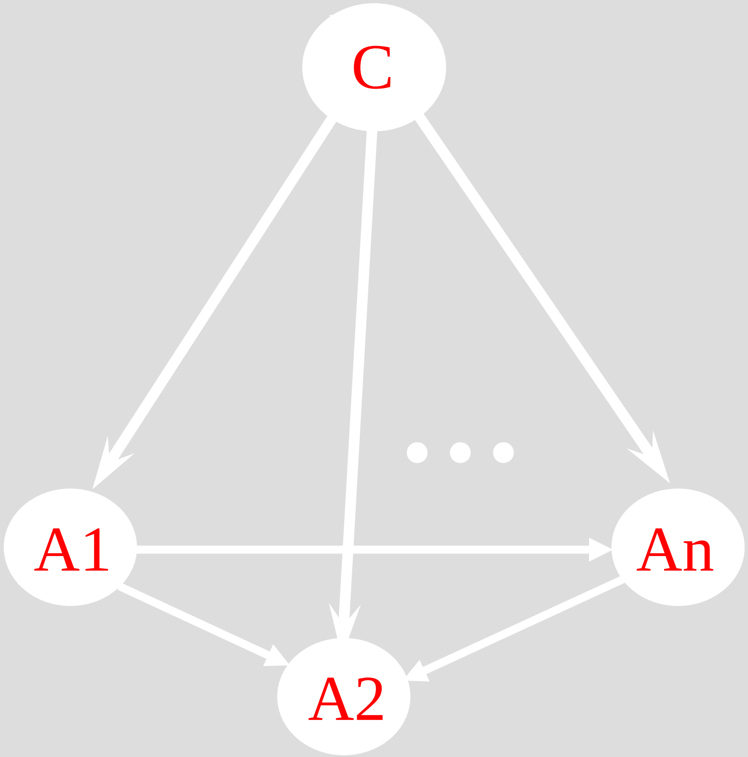
$$P(A_k \mid C_i)$$

Bayesian classifier - Extensions

- TAN



- BAN



Example

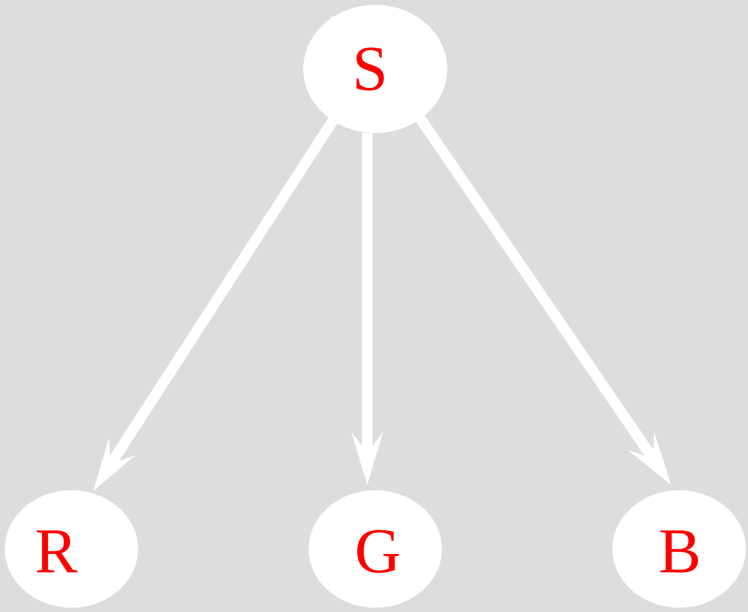
- Skin classification based on color
  - Hypothesis: skin, no-skin
  - Attributes: red, green, blue (256 values each)

- Probability function:

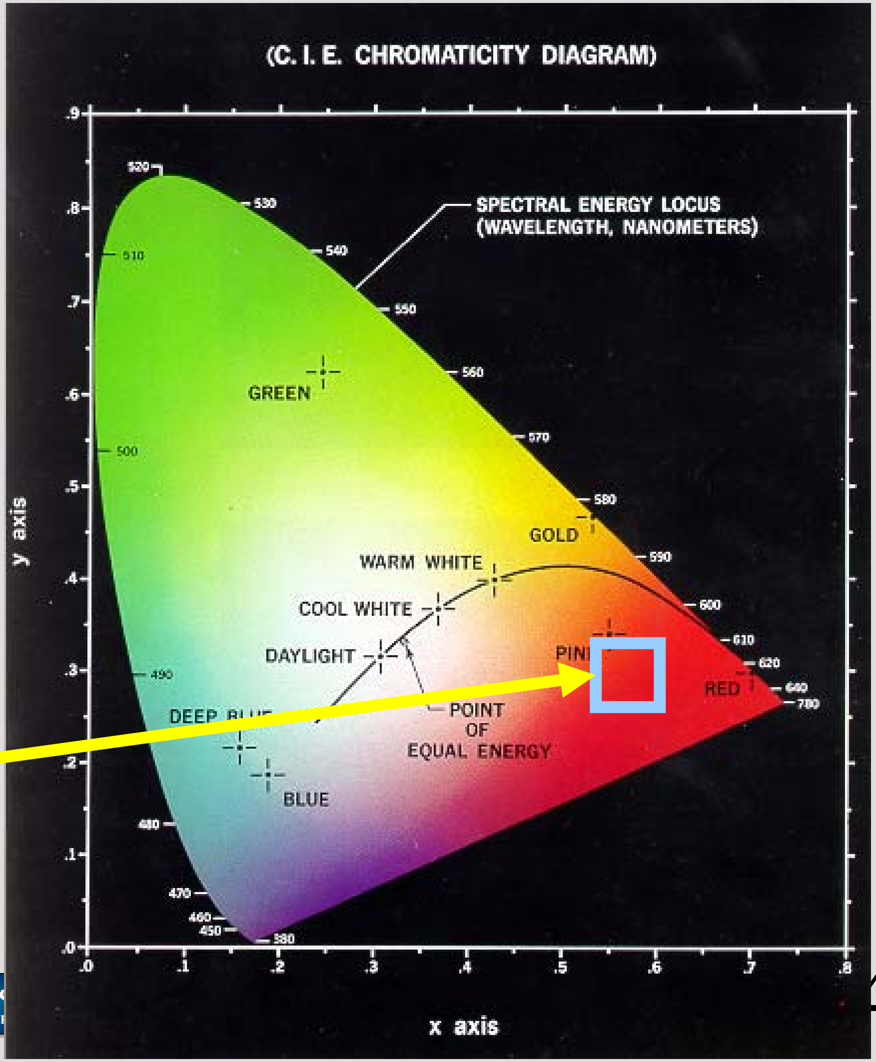
  $P(S|R,G,B) = k\, P(S)\, P(R|S)\, P(G|S)\, P(B|S)$

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Naive Bayes



Color based classification



"Skin" region in RGB space

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Skin detection

Detection of skin pixels based on color information and a Bayesian classifier

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Attribute Selection

- When there are many attributes, it can become impractical to include all in the classifier
- Also, redundant attributes (highly dependent), may reduce the accuracy
- A simple way to select relevant attributes is to select only those that provide information on the class, by measuring their mutual information: $I(C,Ax)$
- The attributes with low I are eliminated

Mutual information

- It is a measure of the dependency between a pair of variables given by:

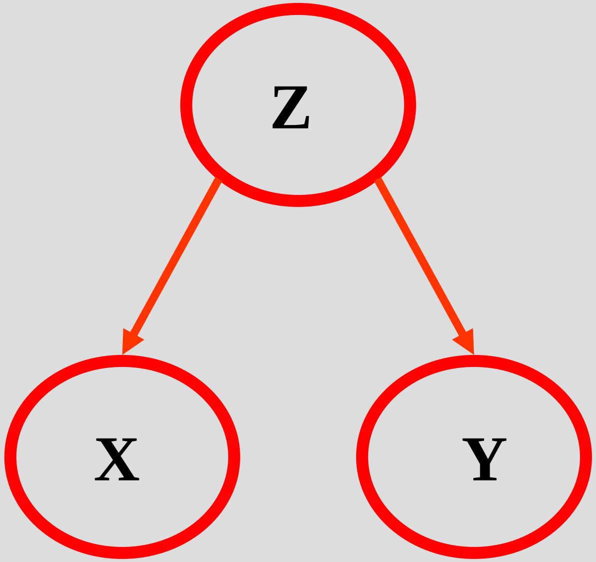
$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i, X_j) \log \frac{P(X_i, X_j)}{P(X_i) P(X_j)}$$

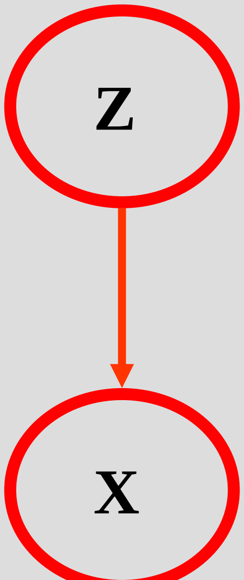
- It can be extended to consider the mutual information of two variables given a third one – conditional mutual information

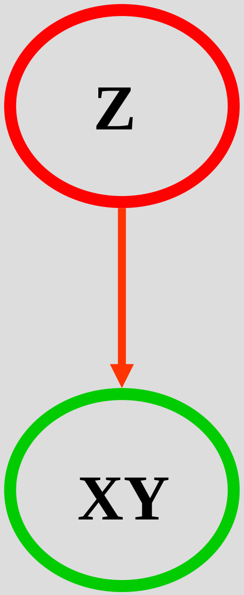
Structural Improvement

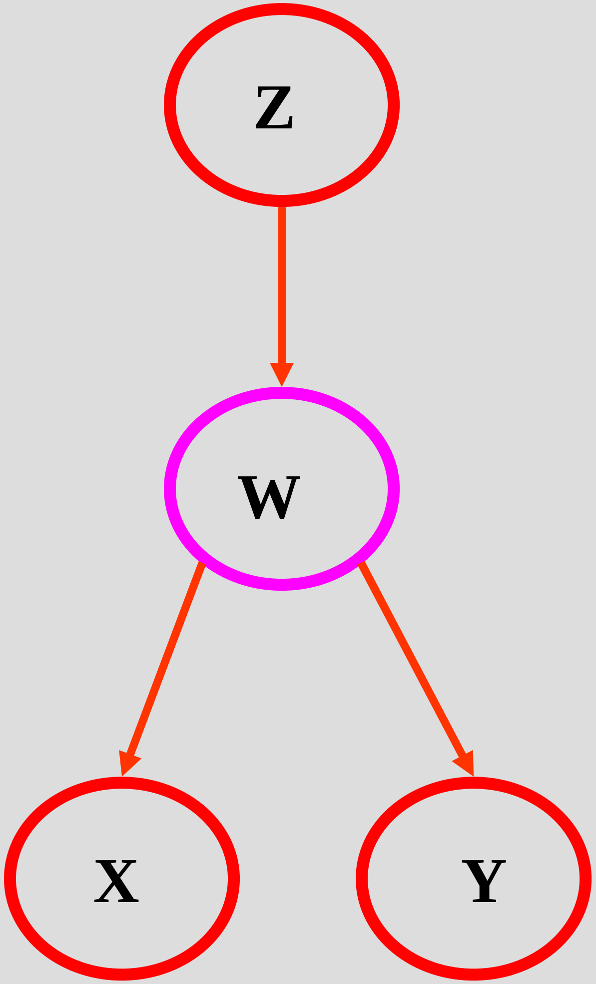
- Start from a subjective structure and improve with data
- Verify conditional independencies:
  - Node elimination
  - Node combination
  - Node insertion

Structural improvement









Olivier.Aycard@imag.fr
esucar@inaoep.mx

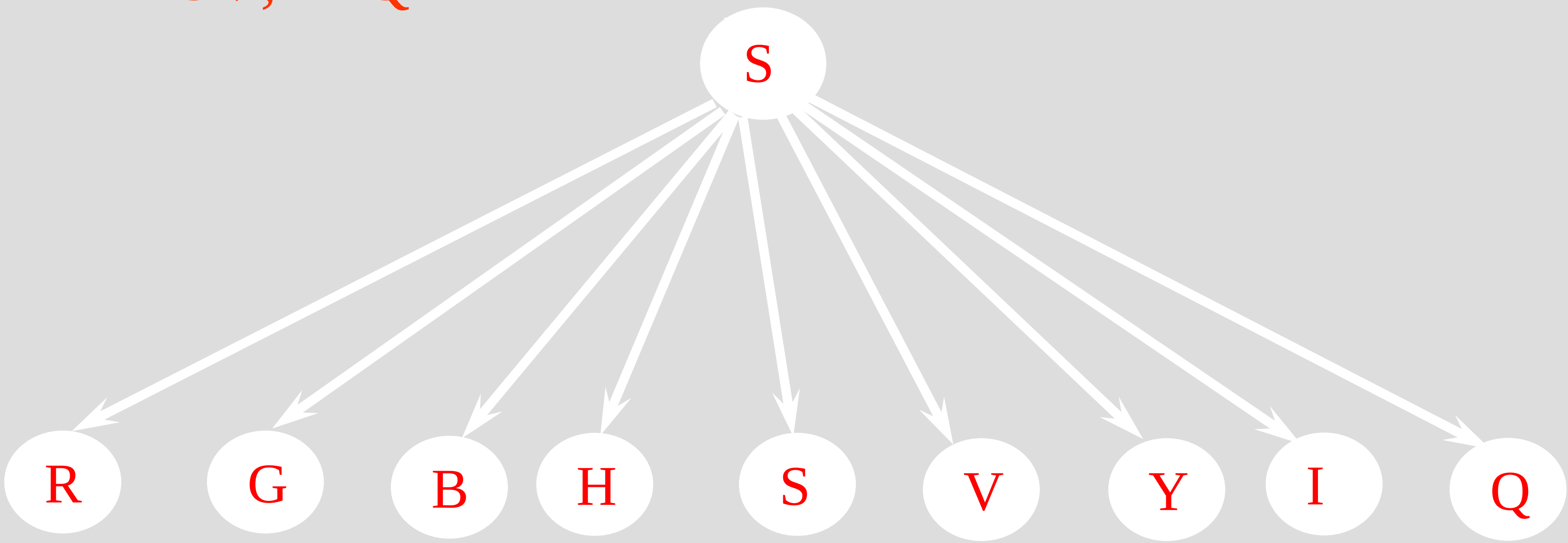
Learning an *optimal* naive Bayes classifier

1. Build an initial classifier with all the attributes
2. Repeat until the classifier can not be improved (based on the MDL principle):
   a. Eliminate redundant attributes
   b. Eliminate/Join dependant attributes
   c. Improve discretization of continuous attributes
3. Test classifier on different data (cross validation)

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Improving skin classification

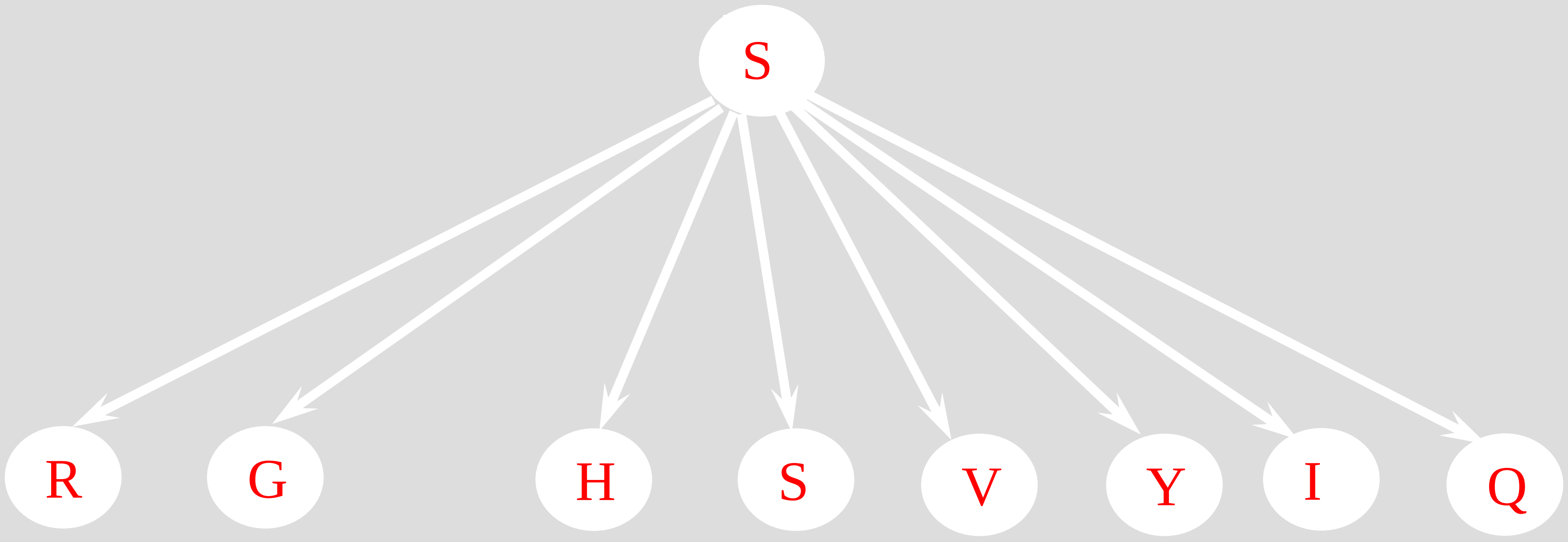
- Nine attributes combining 3 color models: RGB, HSV, YIQ



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structural Improvement

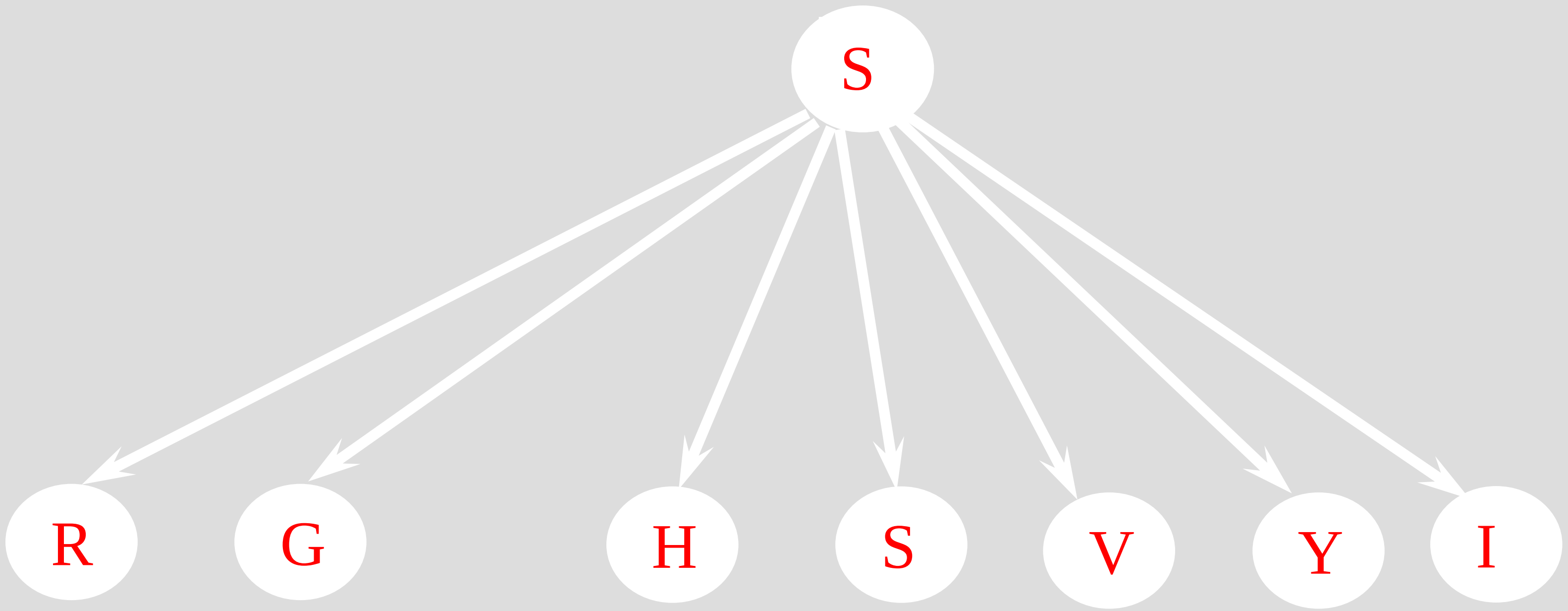
Eliminate B



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structural Improvement

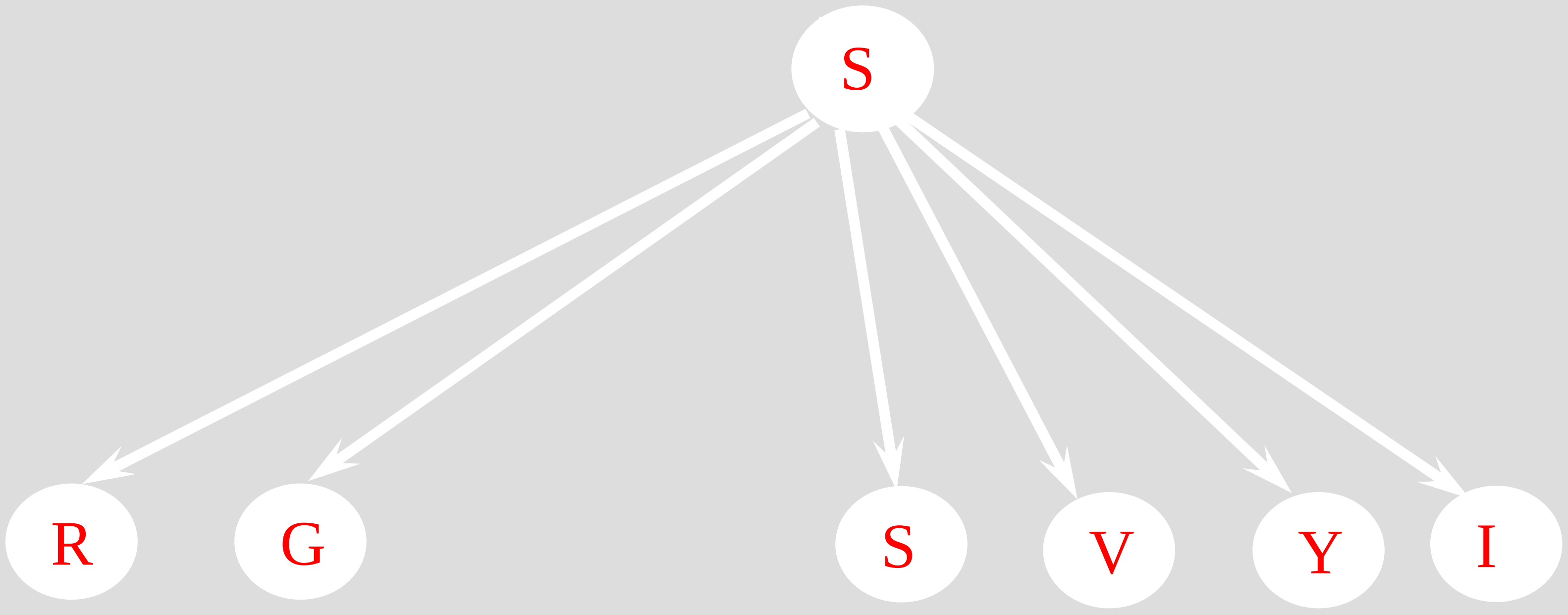
Eliminate Q



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structural Improvement

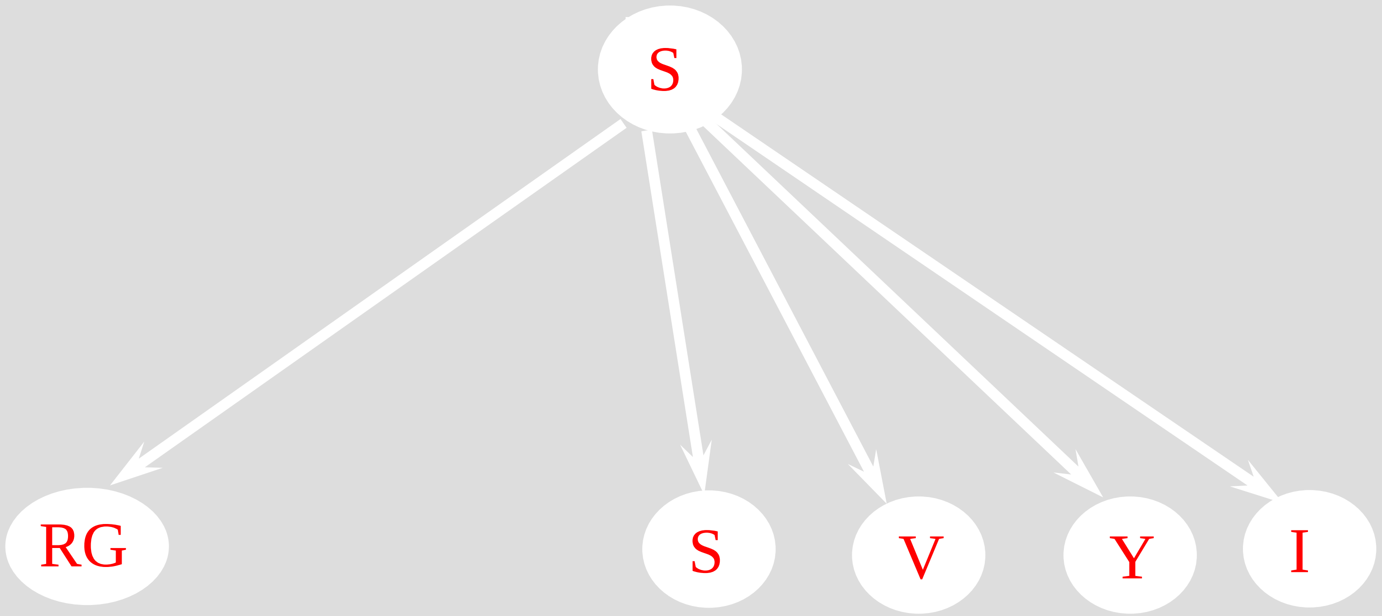
Eliminate H



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structural Improvement

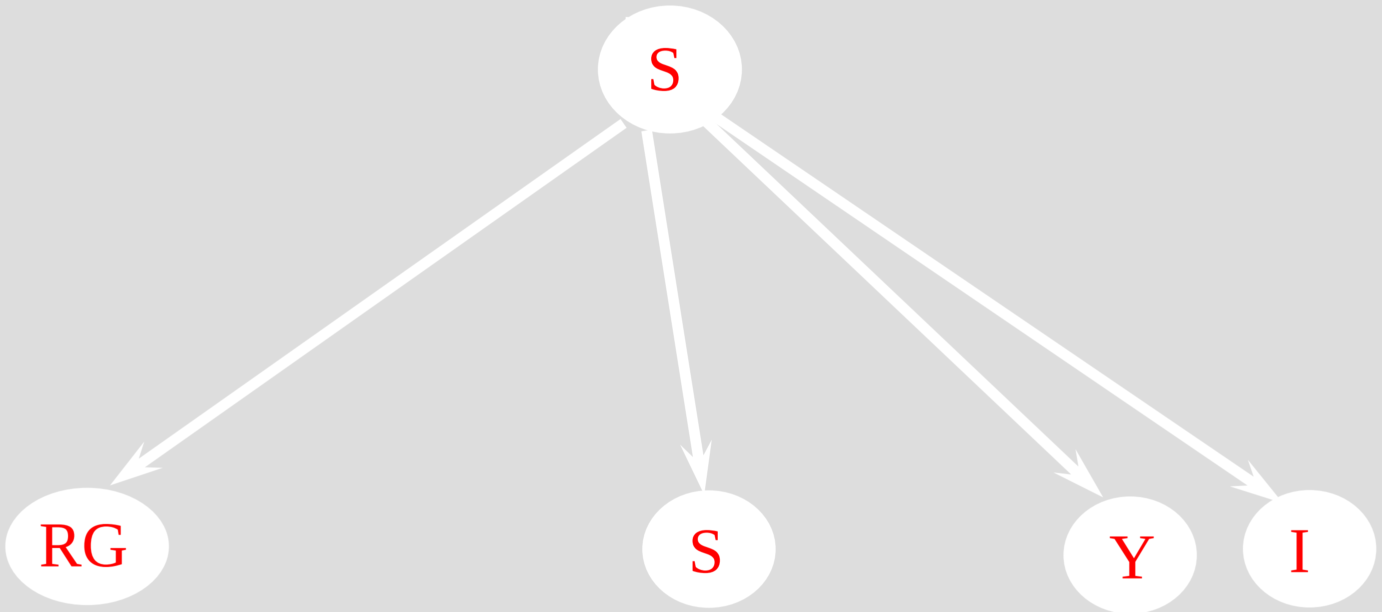
Join RG



Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structural Improvement

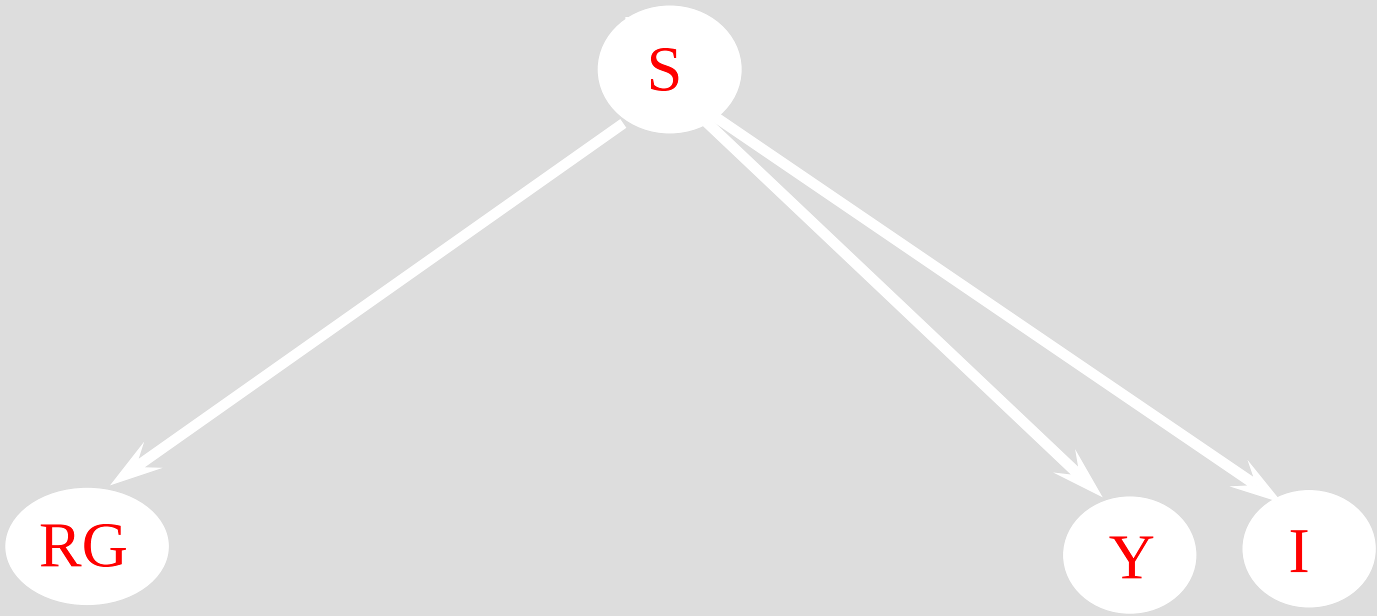
Eliminate V



Structural Improvement

Eliminate S

Acurracy: initial 94%
final 98%



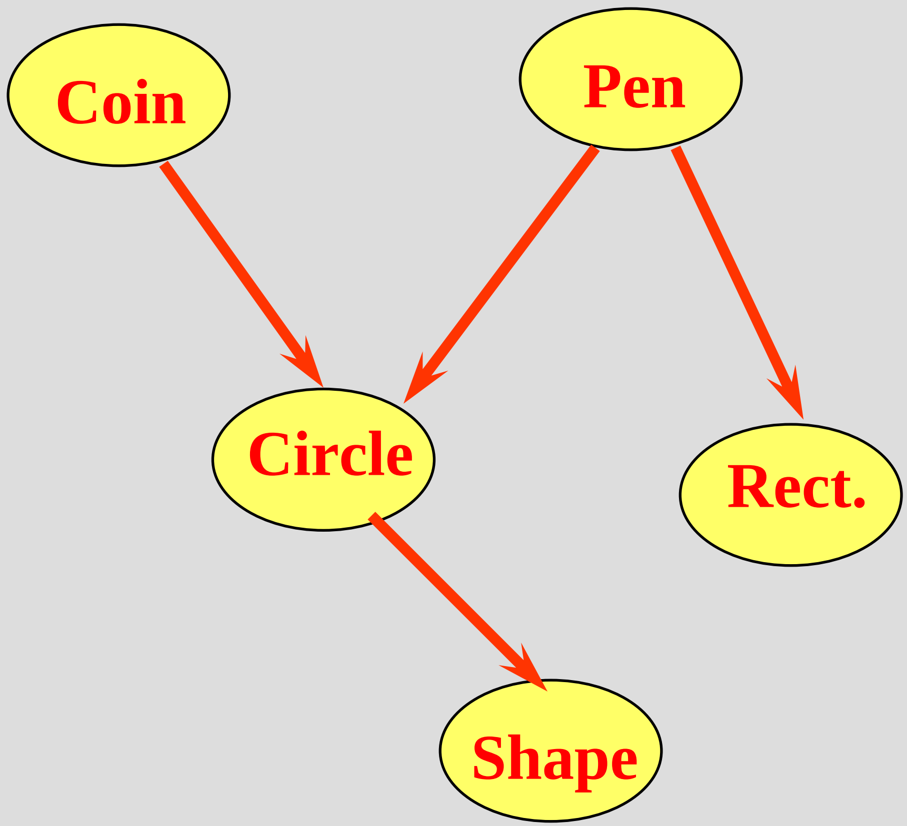
Contents

- Fundamentals of Bayesian Techniques
  - Introduction
  - Fundamentals
  - Bayesian Classifiers
  - Bayesian Networks

Representation

- Bayesian networks (BN) are a graphical representation of dependencies between a set of random variables. A Bayesian net is a Directed Acyclic Graph (DAG) in which:
  - Node: Propositional variable.
  - Arcs: Probabilistic dependencies.
- An arc between two variables represents a direct dependency, usually interpreted as a *causal* relation.

An example of a BN



Interpretation

- Represents (in a compact way) the joint probability distribution of all the variables

- In the previous example:

P(Co, P, Ci, R, S) =

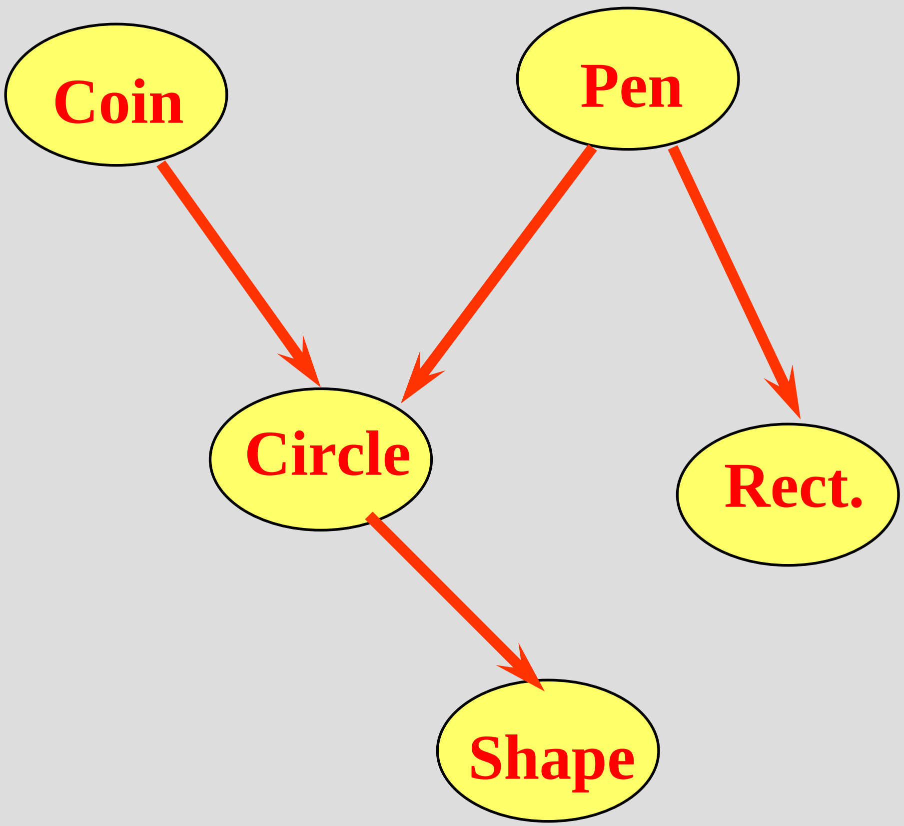
P(Co) P(P) P(Ci|Co,P) P(R|P) P(S|Ci)

Olivier.Aycard@imag.fr
esucar@inaoep.mx

Structure

- The topology of the network represents the dependencies (and independencies) between the variables

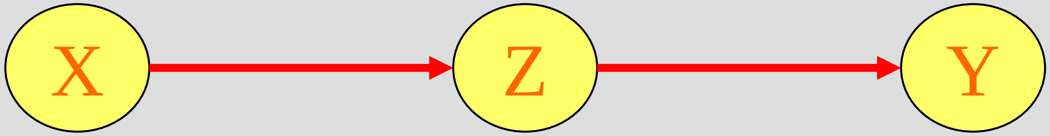
- Conditional independence relations between variables or sets of variables are obtained by a criteria called D-separation



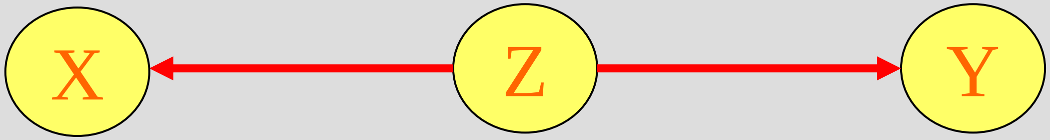
E.g.: {R} *is d-separated from* {Co, Ci, S} *by* {P}

Graphical separation – 3 basic cases

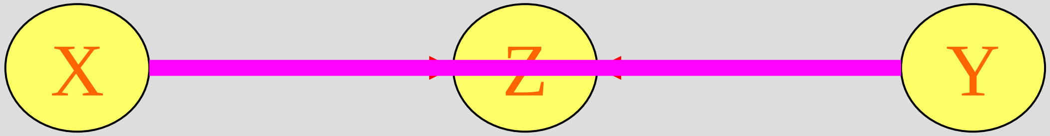
- “Markov”



- “common cause”



- “explaining away”

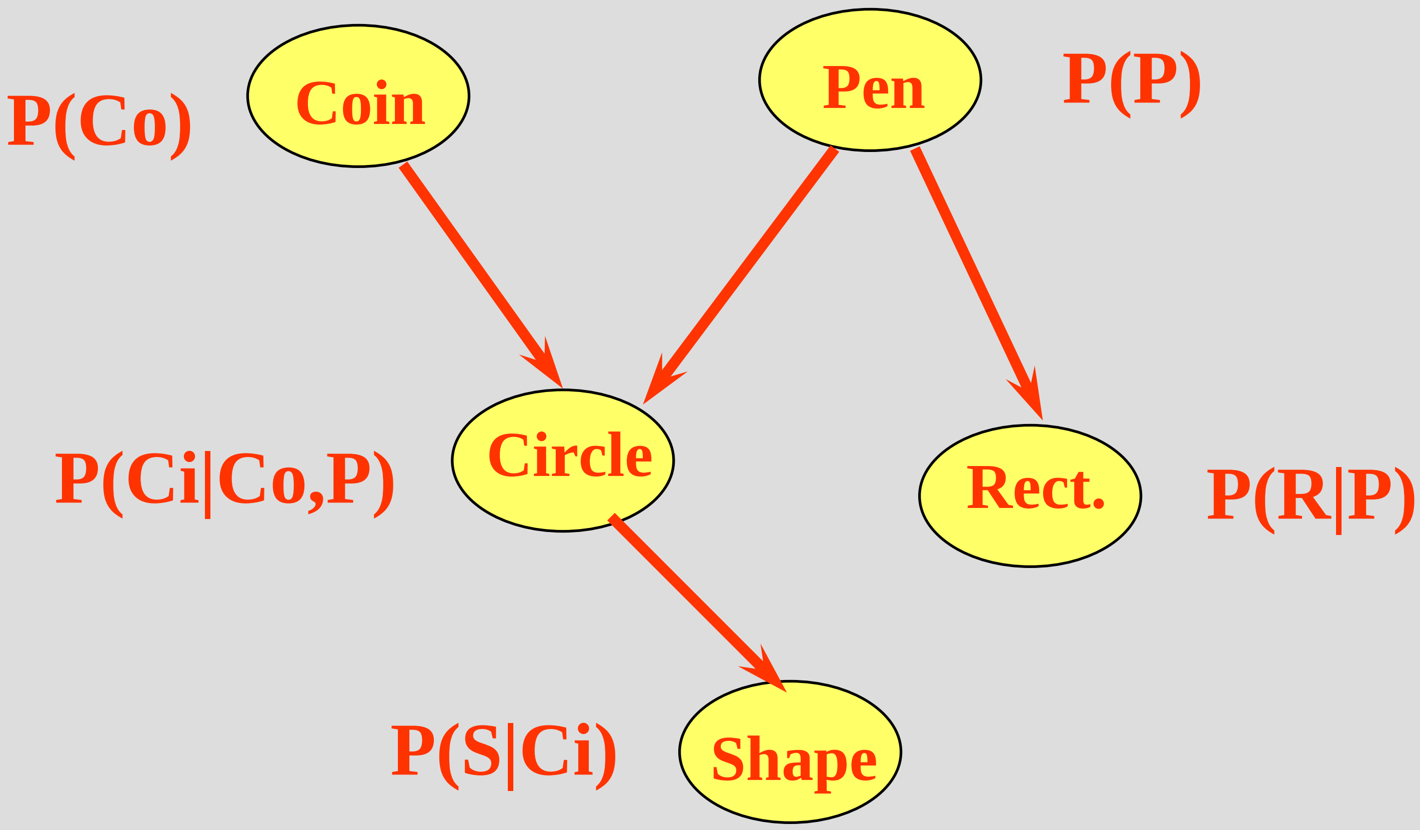


Olivier.Aycard@imag.fr
esucar@inaoep.mx

Parameters

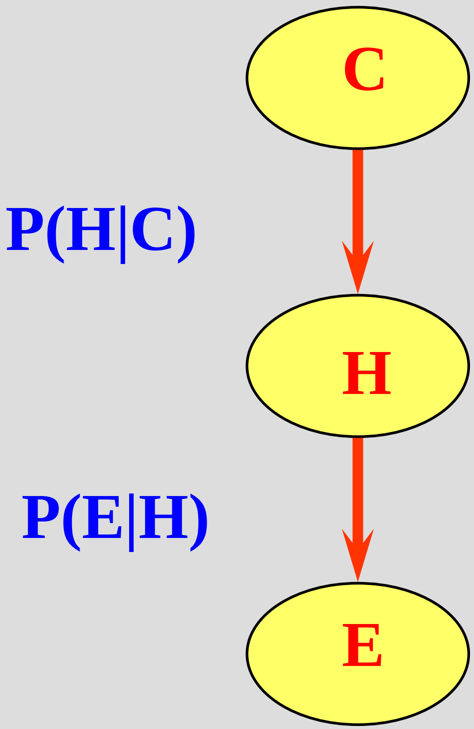
Conditional probabilities of each node given its parents.

- Root nodes: vector of prior probabilities
- Other nodes: matrix of conditional probabilities



P(Co, P, Ci, R, S) =
P(Co) P(P) P(Ci|Co,P) P(R|P) P(S|Ci)

Inference



***Causal*:**

C → H

***Evidential*:**

E → H

***Mixed*:**

C,E → H

Inference

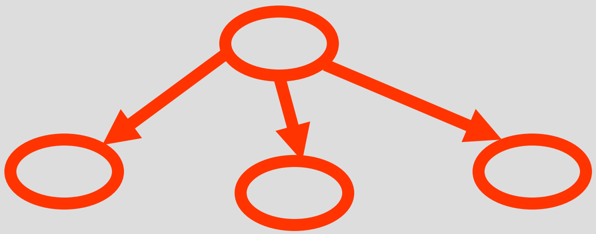
There are several inference algorithms:

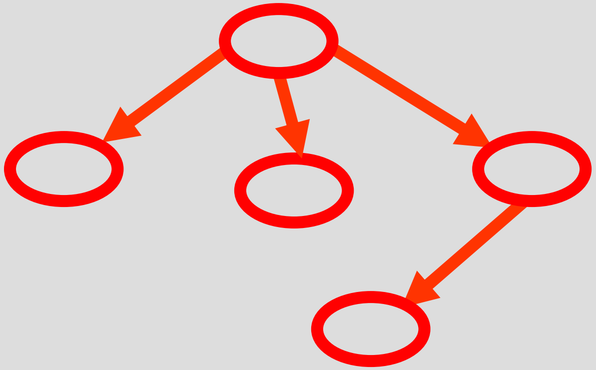
- One variable:
    - Variable elimination
- All the variables:
    - Polytrees:
        - Message passing (Pearl's algorithm)
    - General structure:
        - Junction Tree
        - Stochastic simulation

Types of structures

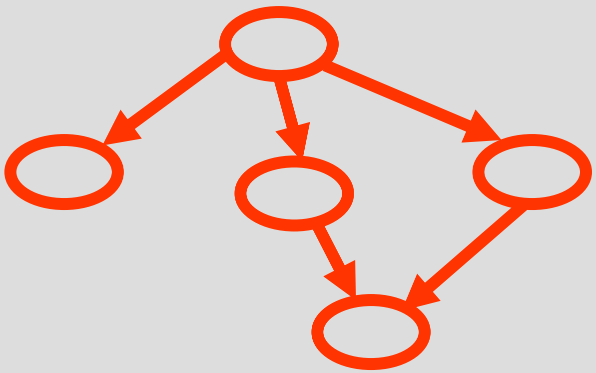
- Trees
- Polytrees

- Singli-connected





- Multiconnected



Olivier.Aycard@imag.fr
esucar@inaoep.mx