Support for Multiple Cause Diagnosis with Bayesian Networks

Presentation
Friday 4 October
Randy Jagt

TU Delft
Technische Universiteit Delft

Decision Systems Laboratory
Intelligent Systems Program
University of Pittsburgh

UNIVERSITY OF PITTSBURGH
VERITAS VIRTUS 1787

Overview

- Introduction
- Objectives
- Research
- Implementation
- GeNIe DIAG
- Tests & Results
- Conclusions

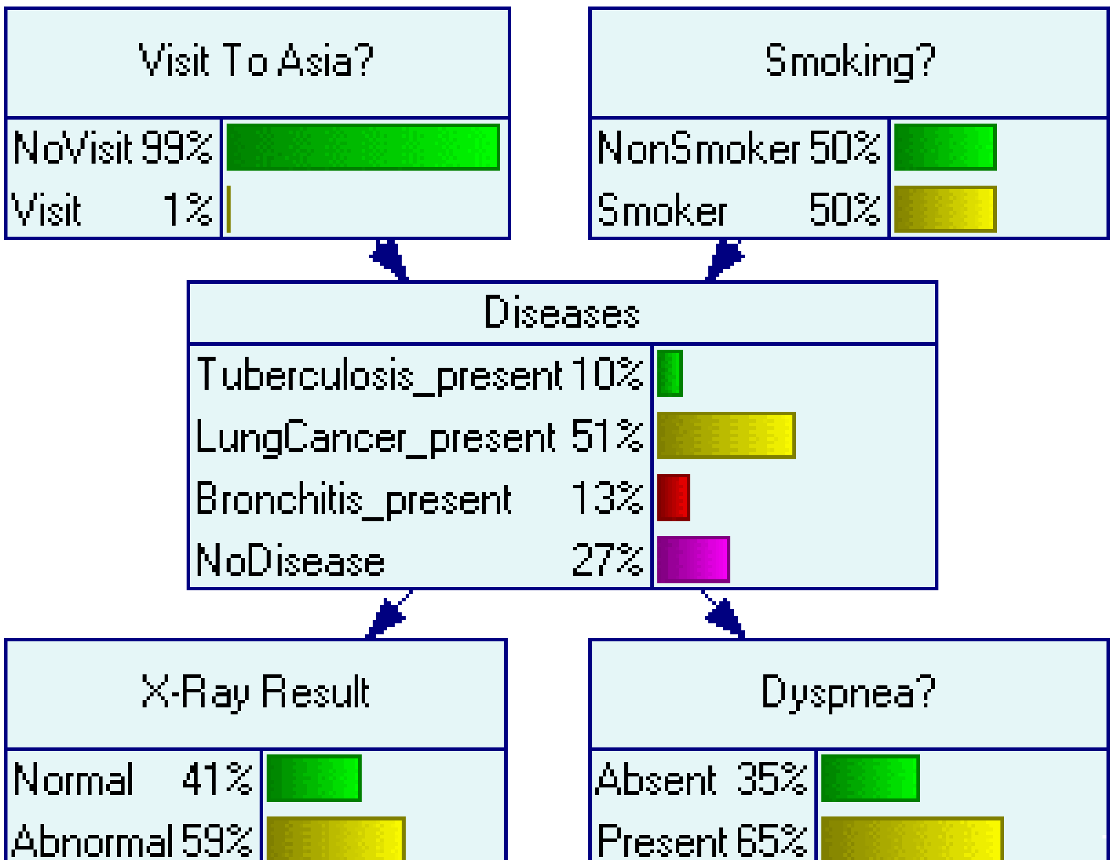
Diagnosis

The process of determining the cause or malfunction by means of collecting information

Diagnostic Expert Systems

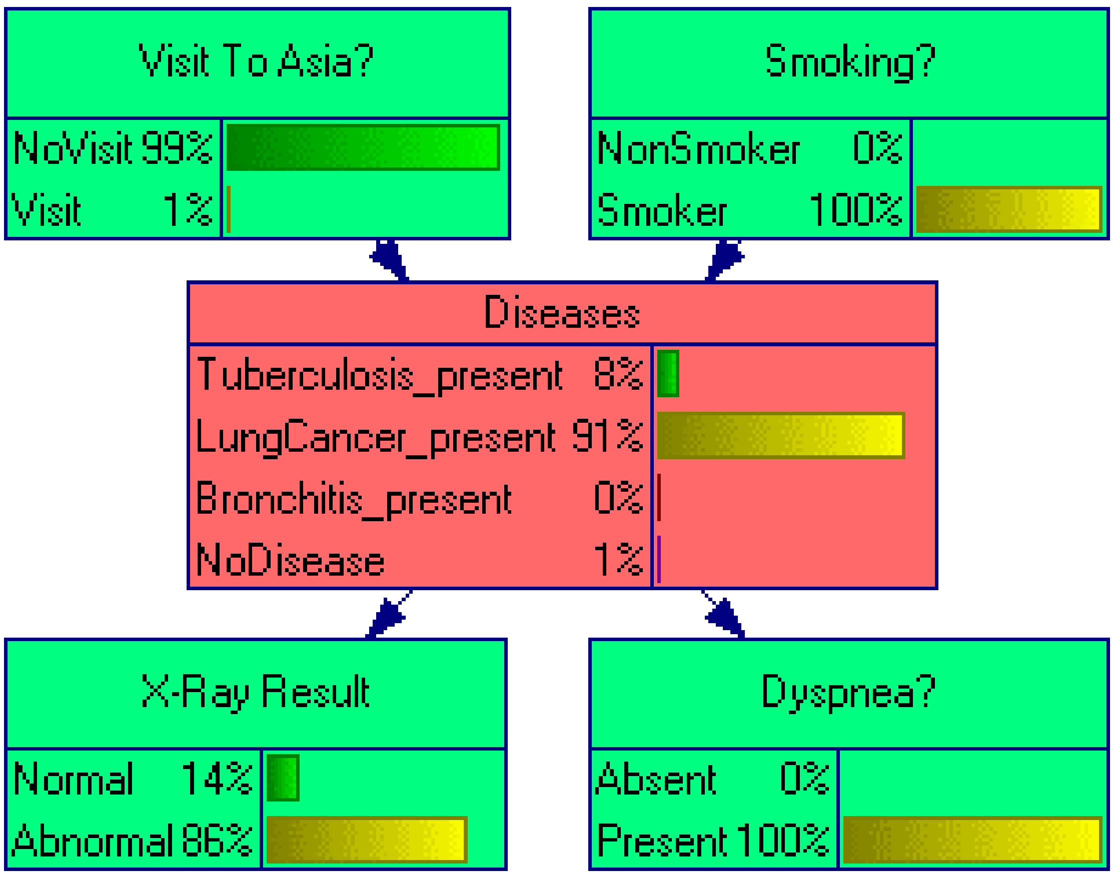
- Primary tasks:
  - *Determine the most probable cause*
  - *Determine which information to gather*
- Applications
  - Medicine
  - Troubleshooting

Bayesian Netwerk



- Random variables
- Probabilistic relations
- Conditional probability table
- Marginal probability distribution
- Reasoning algorithms

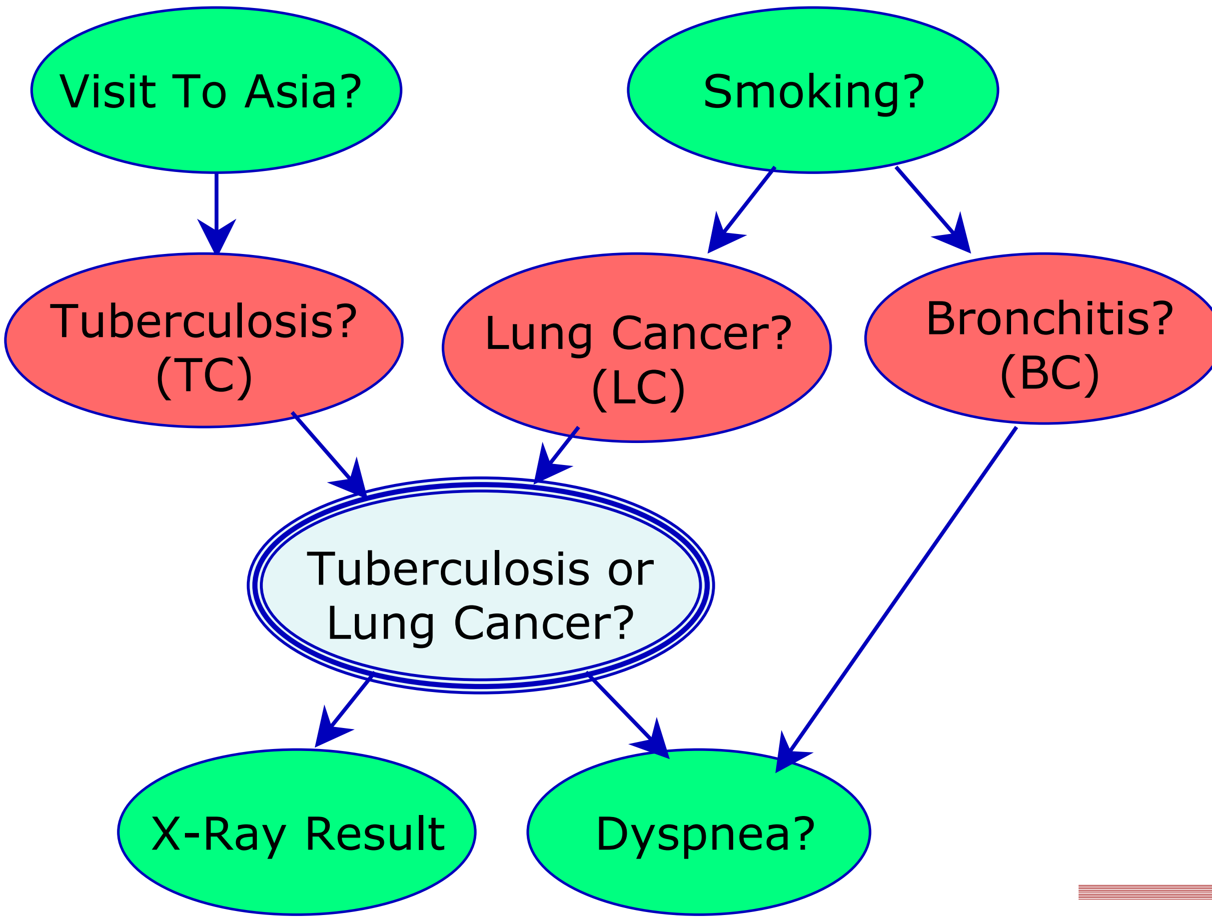
Diagnostis Proces with BN



Test Rankings

Smoking?	0.26
X-Ray Result	0.26
Dyspnea	0.23
Visit To Asia?	<0.01

Suppose Multiple Causes



TC= Tuberculosis
LC= LungCancer
BC= Bronchitis
PN= Pneomia

Problem Statement

- Diagnosing multiple causes leads to the following problems:
  - *Presenting the combinations:*

    User has to keep track on the change of exponential amount of combinations
  - *Calculational effort:*

    To determine the test rankings, the joint probability distribution over all combinations has to be calculated

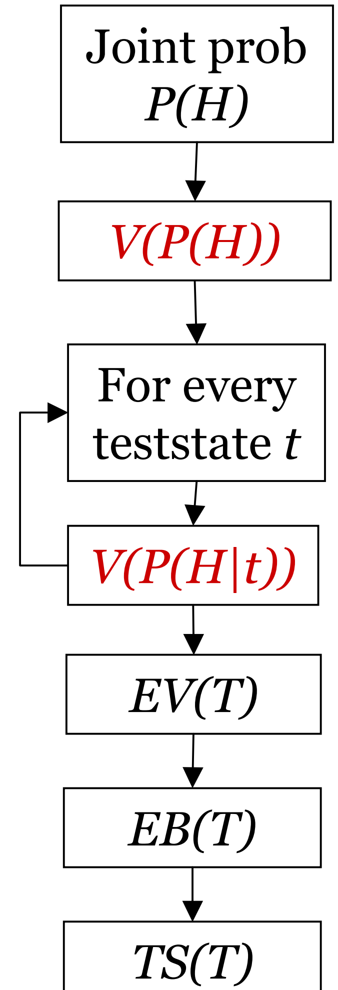
Objectives

- Investigation of Value of Information
- Development of approaches which:
  - Handle the computational complexity
  - Allow the user to work with multiple causes
- Implementation of these approaches
- Testing these approaches

Value of Information

- *Value* functions
  - Functions that assign a value to a distribution
- Applied
  - without a test $V(P(H))$
  - with a test $V(P(H|T))$

VOI process

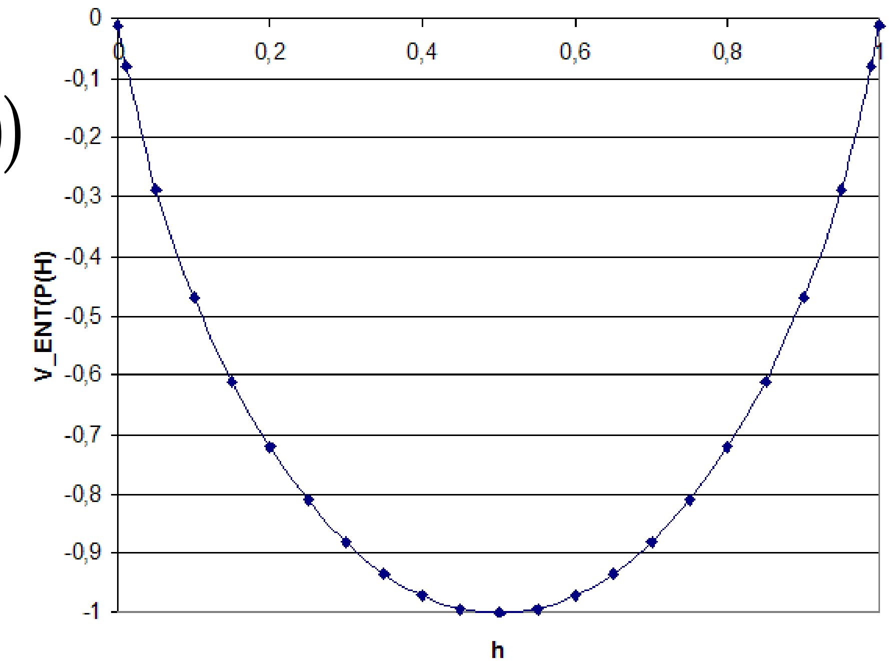


Value Function

- Entropy (Shannon)

$$V_{ENT}(P(H)) = \sum_{h \in H} P(h)\log_2(P(h))$$

  - Minimum at uniform probability
  - Monotonic decreasing function of the number of entries



Value of Information(2)

- Expected Value:

$$EV(T)=\sum_{t\in T} V\big(P(H\mid t)\big)P(t)$$

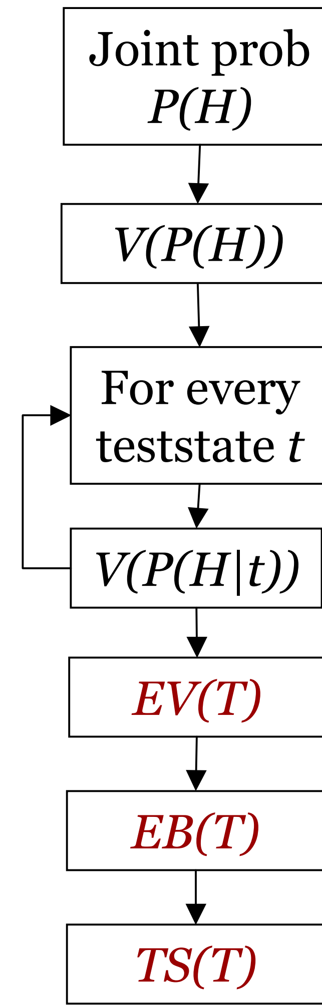
- Expected Benefit:

$$EB(T)=EV(T)-V\big(P(H)\big)$$

- Test Strength:

$$TS=\frac{EB(T)}{V\big(P(H)\big)}$$

VOI process



Marginal Probability Approach

Create new value functions that work with the marginal probability distribution

Marginal & Joint Relation

- Lower bound:

$$P(a,b) \geq \max\{0; P(a)+P(b)-1\}$$

- Upper bound:

$$P(a,b) \leq \min\big(P(a), P(b)\big)$$

- High joint probability ⇔ All high marginal probabilities
- Low joint probability ⇔ At least one low marginal probability

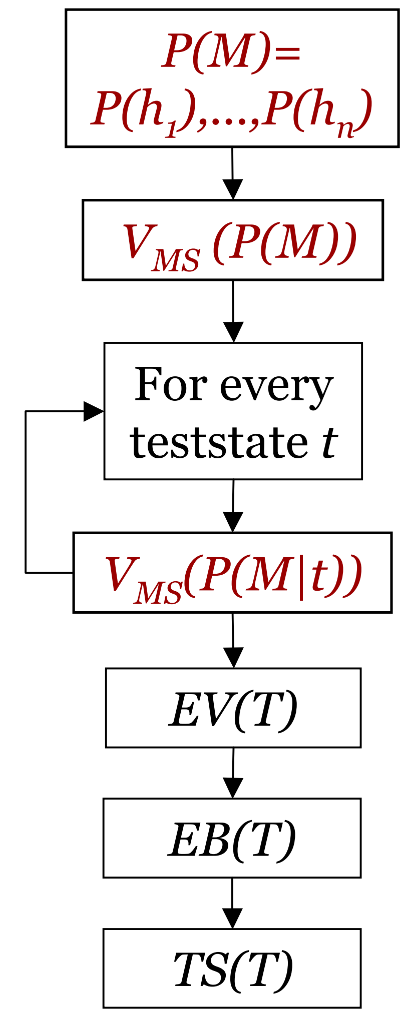
Marginal VOI Process

- Marginal Strength Function

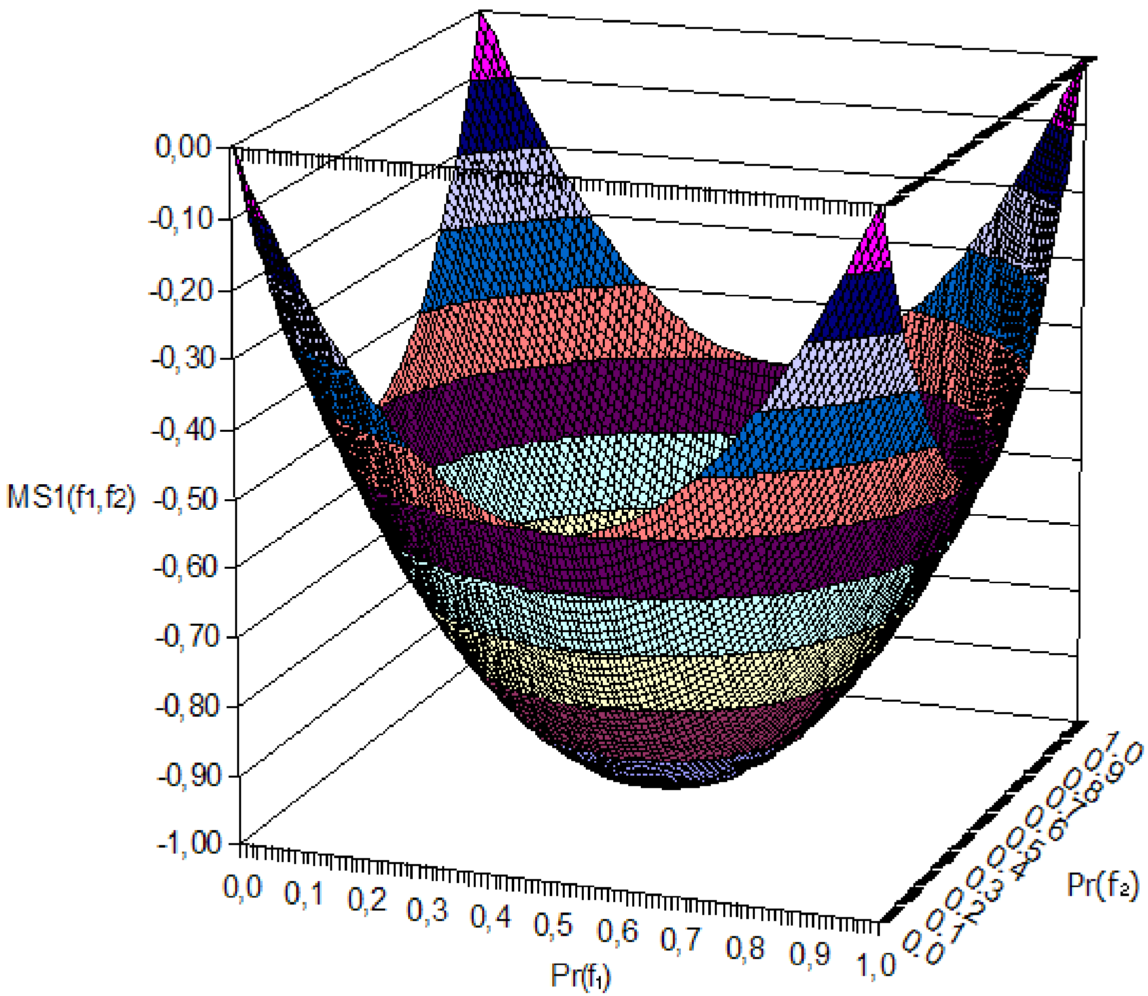
$$V_{MS}(P(M)) = \left(\frac{\sum_{h \in M} (P(h) - 0.5)^2}{\left(\frac{1}{2}\right)^2} - n \right) \cdot \frac{1}{n}$$

  - Minimum at probability 0.5

VOI process



Marginal Strength



Joint Probability Approach

Use of a marginal based function or Copula function to calculate the joint probability distribution

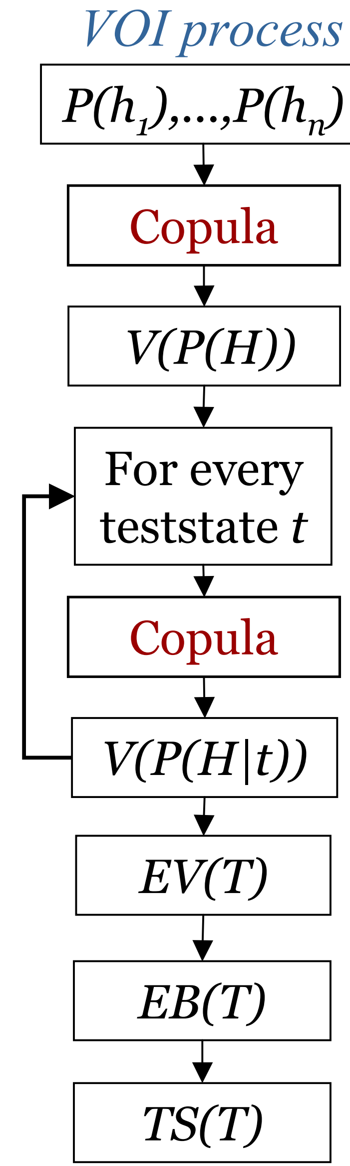
Joint VOI Process

- Product Copula

$$P(h_1,\ldots,h_n)=\prod_{i=1}^{n}P(h_i)$$

- Entropy

$$V_{ENT}(P(H))=\sum_{h\in H}P(h)\log_2(P(h))$$



Presentational Aspect

- Denote interesting states as targets

Tuberculosis?
Absent
Present

LungCancer?
Absent
Present

Bronchitis?
Absent
Present

- Let the user choose targets to pursue

Target states
TC_present
LC_present
BC_present

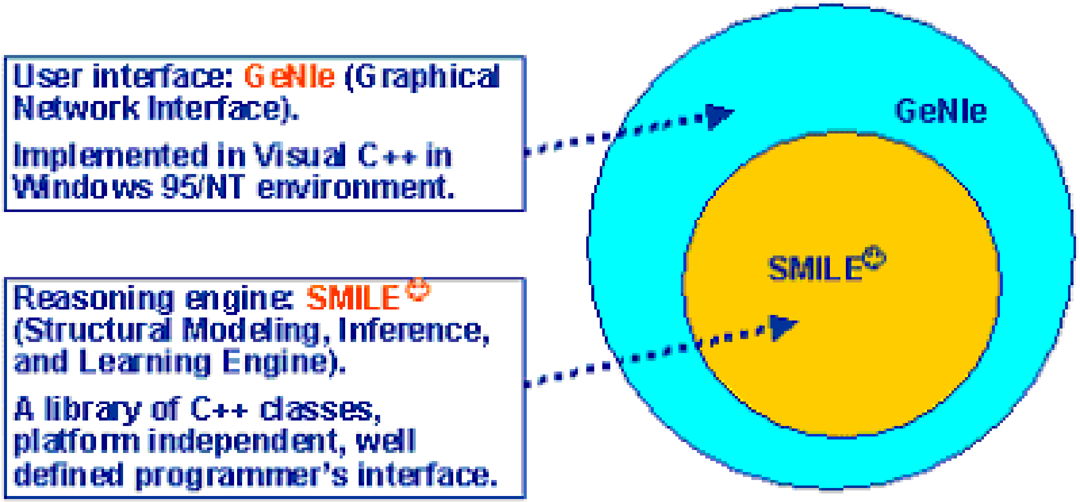
Diagnosis with Targets

- Marginal probability approach over all pursued targets
- Joint probability approach over all the combinations with at least one pursued target and the rest

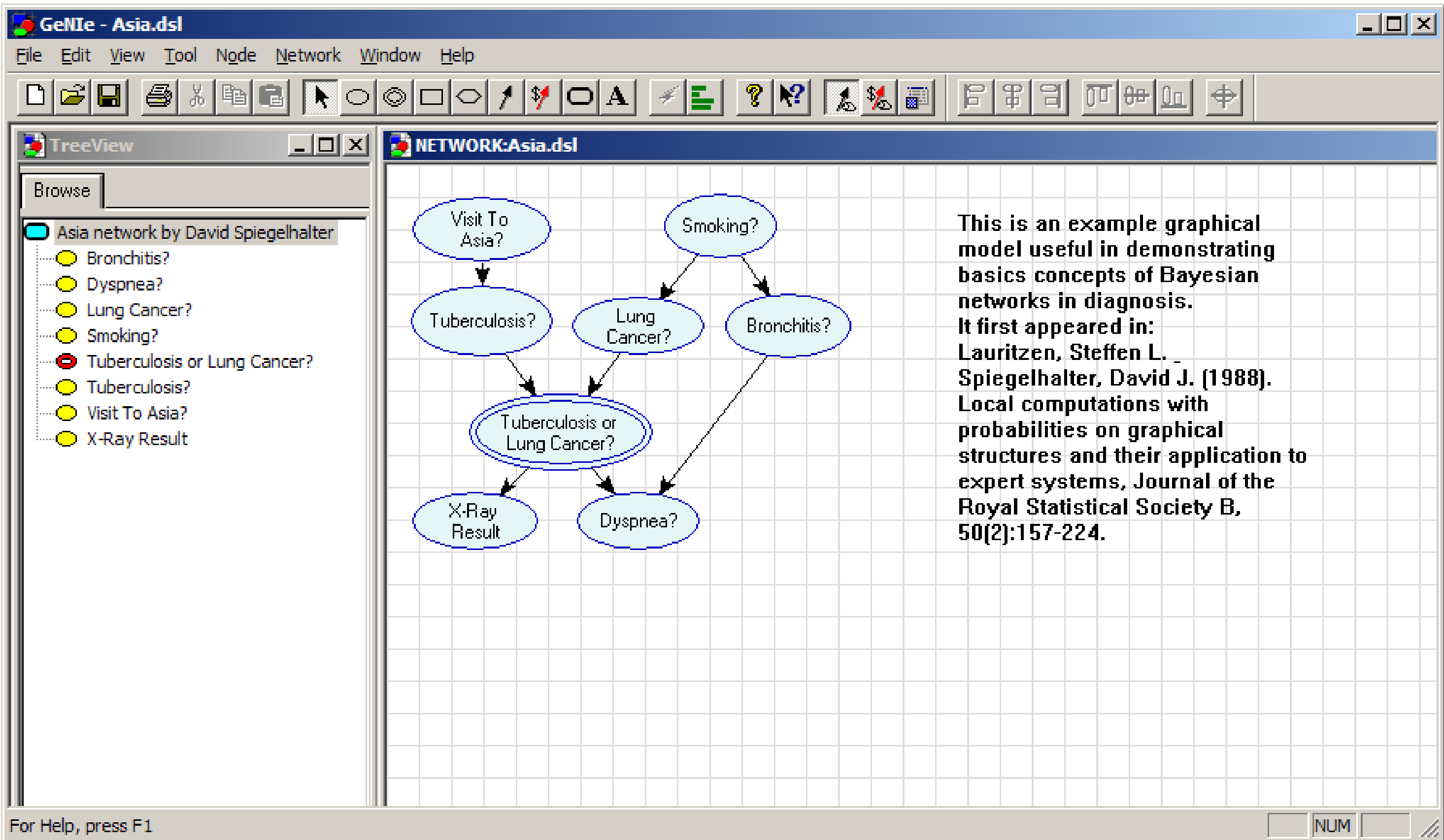
TC_absent LC_absent BC_absent	TC_present LC_absent BC_absent	TC_absent LC_absent BC_present	TC_present LC_absent BC_present
TC_absent LC_present BC_absent	TC_present LC_present BC_absent	TC_absent LC_present BC_present	TC_present LC_present BC_present

Implementation

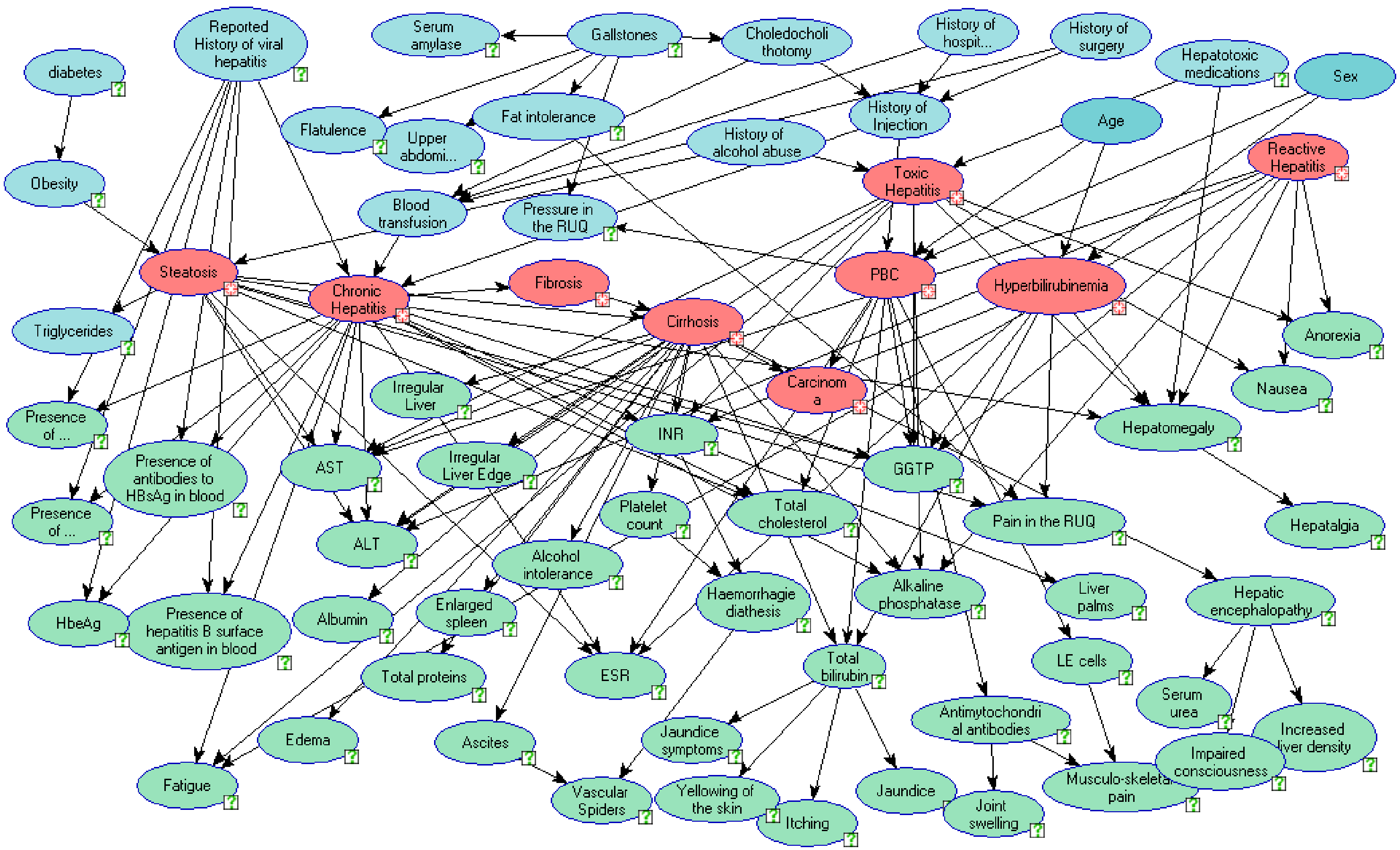
- SMILE (Structural Modeling, Inference, and Learning Engine)
- GeNIe (Graphical Network Interface)
- Visual C++
- GeNIe DIAG



GeNIe DIAG


GeNIe - Asia.dsl
File Edit View Tool Node Network Window Help
TreeView
Browse
Asia network by David Spiegelhalter
Bronchitis?
Dyspnea?
Lung Cancer?
Smoking?
Tuberculosis or Lung Cancer?
Tuberculosis?
Visit To Asia?
X-Ray Result
NETWORK:Asia.dsl
Visit To Asia?
Smoking?
Tuberculosis?
Lung Cancer?
Bronchitis?
Tuberculosis or Lung Cancer?
X-Ray Result
Dyspnea?
This is an example graphical model useful in demonstrating basics concepts of Bayesian networks in diagnosis.
It first appeared in:
Lauritzen, Steffen L. _ Spiegelhalter, David J. (1988). Local computations with probabilities on graphical structures and their application to expert systems, Journal of the Royal Statistical Society B, 50(2):157-224.
For Help, press F1
NUM

Time Test



Reported History of viral hepatitis
Serum amylase
Gallstones
Choledocholithotomy
History of hospit...
History of surgery
Hepatotoxic medications
Sex
diabetes
Flatulence
Upper abdomi...
Fat intolerance
History of Injection
Age
History of alcohol abuse
Obesity
Blood transfusion
Pressure in the RUQ
Toxic Hepatitis
Reactive Hepatitis
Steatosis
Chronic Hepatitis
Fibrosis
PBC
Hyperbilirubinemia
Cirrhosis
Triglycerides
Anorexia
Irregular Liver
Carcinoma
Nausea
Presence of ...
INR
Hepatomegaly
Presence of antibodies to HBsAg in blood
AST
Irregular Liver Edge
GGTP
Presence of ...
Platelet count
Total cholesterol
Pain in the RUQ
Hepatalgia
ALT
Alcohol intolerance
Alkaline phosphatase
Liver palms
Hepatic encephalopathy
HbeAg
Presence of hepatitis B surface antigen in blood
Albumin
Enlarged spleen
Haemorrhagic diathesis
LE cells
Total proteins
ESR
Total bilirubin
Serum urea
Antimytochondrial antibodies
Increased liver density
Edema
Ascites
Jaundice symptoms
Impaired consciousness
Musculo-skeletal pain
Fatigue
Vascular Spiders
Yellowing of the skin
Itching
Jaundice
Joint swelling

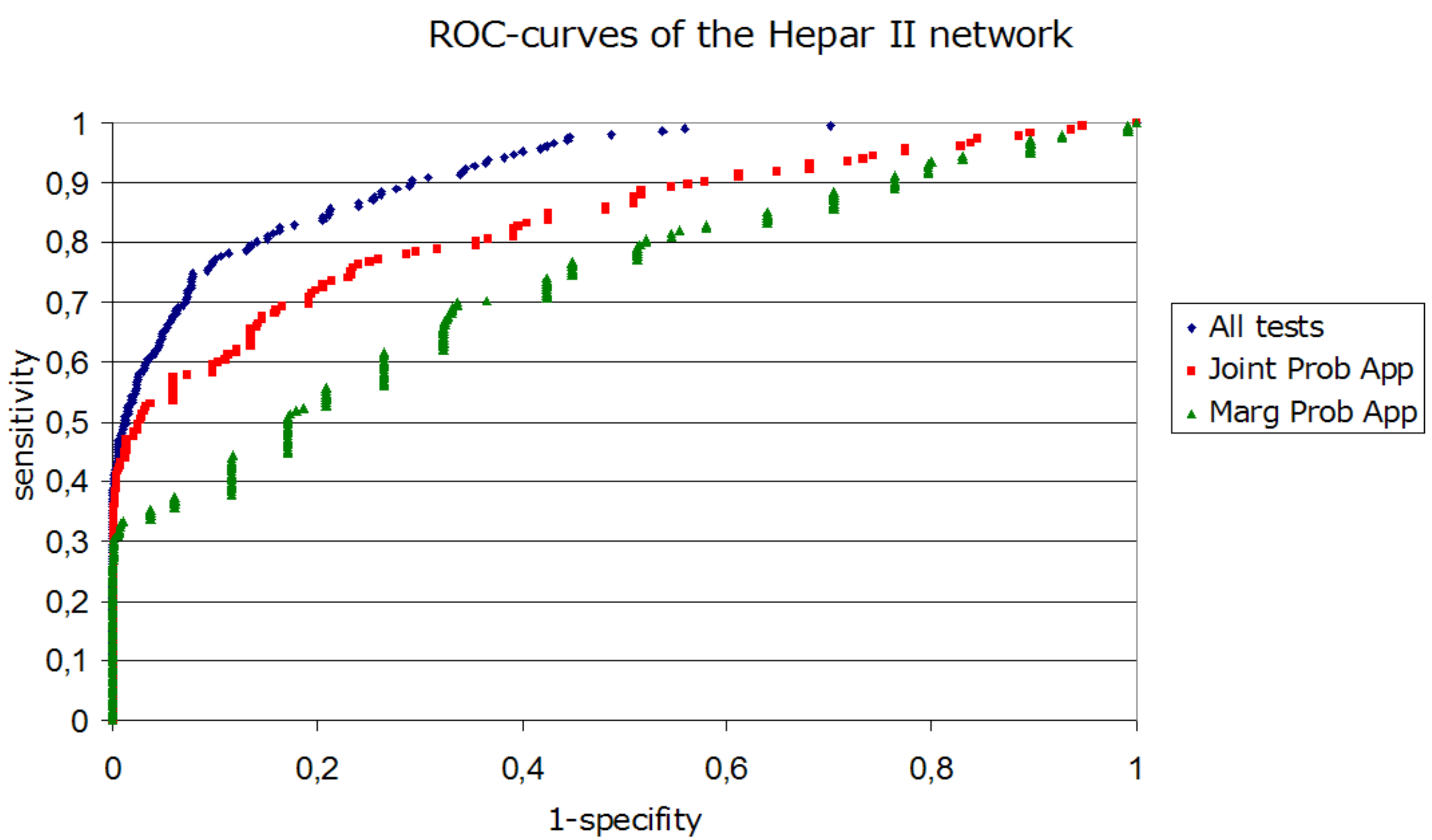
Time Test Results

Time in Seconds	Asia network 3 targets	Hepar II network 10 targets	Pitt network 47 targets
Marginal Probabiltiy Approach	0	0	0
Joint Probability Approach	0	1	>60 min
True Joint Probability	0	514	>60 min

Quality Test & ROC-Analysis

- ROC-curve
  - Sensitivity = likelihood that a present cause was correctly diagnosed
  - Specifity = likelihood that an absence of cause was correctly diagnosed
- Hepar II network

Quality Test Results



Conclusions

- Development of two approaches
- Implemented in GeNIe & SMILE
- Ability to direct the diagnostic process
- Tests showed that
  - Marginal probability approach=fast but less qaulitative
  - Joint probability approach= slow but good quality

Future Research

- Other copula functions
- Smart algorithm for calculating the true joint probability distribution
- Expansion of the multiple cause support to multiple test ranking

Questions?