BCCS 2008/09: GM&CSS

Lecture 6:

Bayes(ian) Net(work)s and Probabilistic Expert Systems

A. Motivating examples

- Forensic genetics
- Expert systems in medical and engineering diagnosis
- Bayesian hierarchical models
- Simple applications of Bayes' theorem
- Markov chains and random walks

The 'Asia' (chest-clinic) example

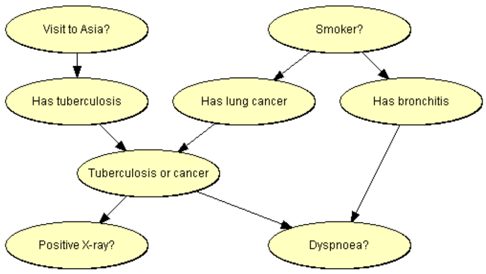
Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, more than one of these diseases or none of them.

A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis.

The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.

+2

Visual representation of the Asia example - a graphical model



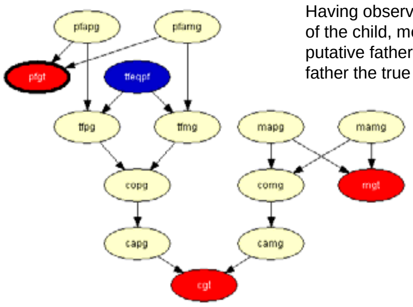
The 'Asia' (chest-clinic) example

Now … a patient presents with shortness-of-breath (dyspnoea) …. How can the physician use available tests (X-ray) and enquiries about the patient's history (smoking, visits to Asia) to help to diagnose which, if any, of tuberculosis, lung cancer, or bronchitis is the patient probably suffering from?

An example from forensic genetics

DNA profiling based on STR's (single tandem repeats) are finding many uses in forensics, for identifying suspects, deciding paternity, etc. Can we use Mendelian genetics and Bayes' theorem to make probabilistic inference in such cases?

Graphical model for a paternity enquiry - allowing mutation



Having observed the genotype of the child, mother and putative father, is the putative father the true father?

Surgical rankings

- 12 hospitals carry out different numbers of a certain type of operation:
  47, 148, 119, 810, 211, 196, 148, 215, 207, 97, 256, 360 respectively.
- They are differently successful, and there are:
  0, 18, 8, 46, 8, 13, 9, 31, 14, 8, 29, 24 fatalities, respectively.

Surgical rankings, continued

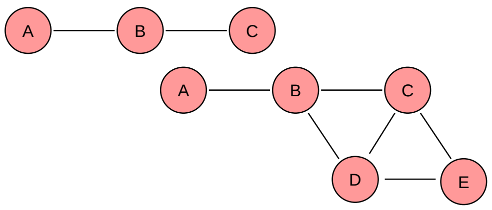
- What inference can we draw about the relative qualities of the hospitals based on these data?
- Does knowing the mortality at one hospital tell us anything at all about the other hospitals - that is, can we 'pool' information?

B. Key ideas in exact probability calculation in complex systems

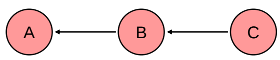
- Graphical model (usually a directed acyclic graph)
- Conditional independence graph
- Decomposability
- Probability propagation: 'message-passing'

Conditional independence graphs

A Conditional independence graph (CIG) has variables as nodes and (undirected) edges between pairs of nodes – <u>absence</u> of an edge between A and C means $A \perp C|$ (rest), e.g.

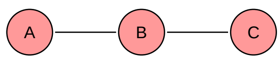


Directed acyclic graph (DAG)



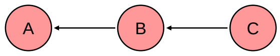
… indicating that model is specified by *p(C)*, *p(B|C)* and *p(A|B)*: $p(A,B,C) = p(A|B)p(B|C)p(C)$

A corresponding Conditional independence graph (CIG) is

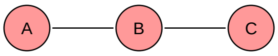


… encoding various conditional independence assumptions, e.g. $p(A,C|B) = p(A|B)p(C|B)$

DAG



CIG



$$p(A,B,C) = p(A,B)p(C\mid A,B) = p(A,B)p(C\mid B)$$

$$= \frac{p(A,B)p(B,C)}{p(B)}$$

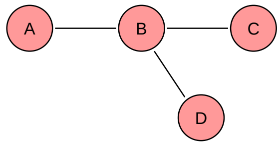
true for any *A, B, C*

definition of *p(C|B)*

since $C \perp A \mid B$

+4

CIG



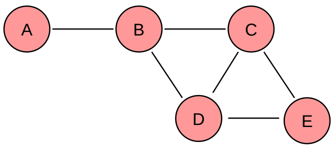
$$p(A,B,C,D) = p(A,B)p(C\mid A,B)p(D\mid A,B,C)$$

$$= p(A,B)p(C\mid B)p(D\mid B)$$

$$= \frac{p(A,B)p(B,C)p(B,D)}{p(B)p(B)}$$

+2

CIG



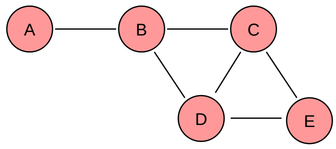
$$p(A,B,C,D,E) = p(A,B)p(C,D\mid A,B)p(E\mid A,B,C,D)$$

$$= p(A,B)p(C,D\mid B)p(E\mid C,D)$$

$$= \frac{p(A,B)p(B,C,D)p(C,D,E)}{p(B)p(C,D)}$$

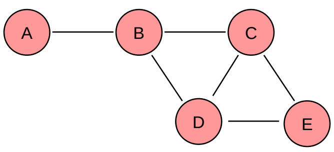
+2

CIG



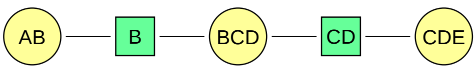
$$p(A,B,C,D,E) = \frac{p(A,B)p(B,C,D)p(C,D,E)}{p(B)p(C,D)}$$

CIG



$$p(A,B,C,D,E) = \frac{p(A,B)p(B,C,D)p(C,D,E)}{p(B)p(C,D)} = \frac{\prod_{\text{cliques}\,C} p(X_C)}{\prod_{\text{separators}\,S} p(X_S)}$$

JT



+1

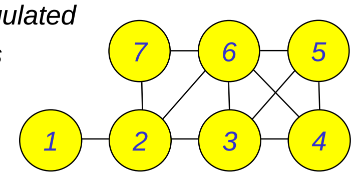
Decomposability

An important concept in processing information through undirected graphs is *decomposability*

(= graph triangulated

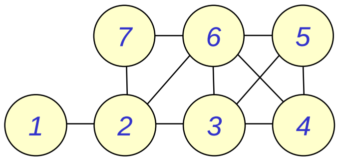
= no chordless

≥ 4-cycles)

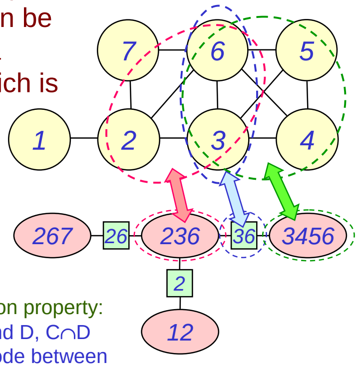


Cliques

A *clique* is a *maximal complete subgraph*:
here the cliques are
{1,2},{2,6,7}, {2,3,6}, and {3,4,5,6}



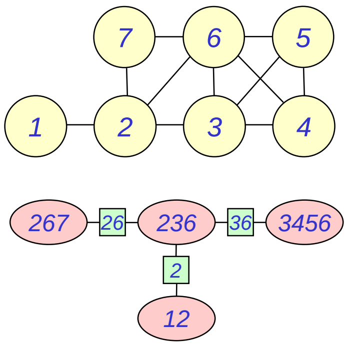
A graph is decomposable if and only if it can be represented by a junction tree (which is not unique)



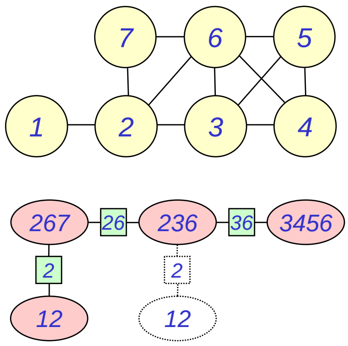
The running intersection property:
For any 2 cliques C and D, C∩D is a subset of every node between them in the junction tree

+4

Non-uniqueness of junction tree



Non-uniqueness of junction tree

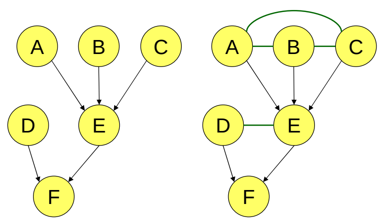


C. Exact probability calculation in complex systems

0. Start with a directed acyclic graph
1. Find corresponding Conditional Independence Graph
2. Ensure decomposability
3. Probability propagation: 'message-passing'

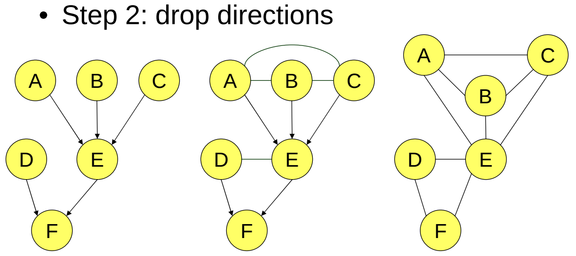
1. Finding an (undirected) conditional independence graph for a given DAG

- Step 1: moralise (parents must marry)

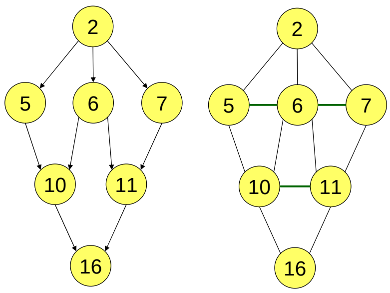


1. Finding an (undirected) conditional independence graph for a given DAG

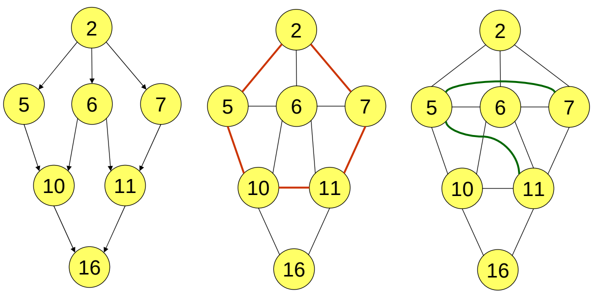
- Step 2: drop directions



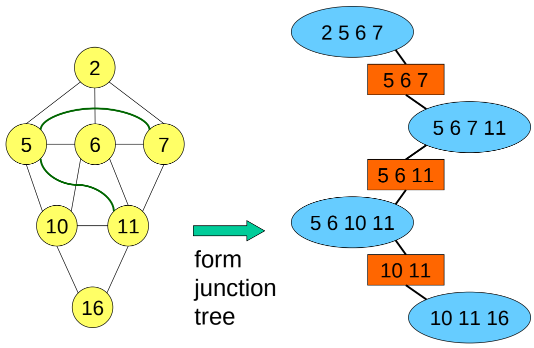
2. Ensuring decomposability



2. Ensuring decomposability
…. triangulate



3. Probability propagation

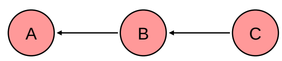


If the distribution *p(X)* has a decomposable CIG, then it can be written in the following potential representation form:

$$p(X) = \frac{\prod_{cliques C} \psi(X_c)}{\prod_{separators S} \psi(X_s)}$$

the individual terms are called potentials; the representation is not unique

DAG



A|B	A=0	A=1
B=0	3/4	1/4
B=1	2/3	1/3

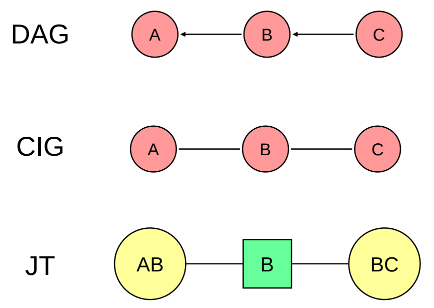
B|C	B=0	B=1
C=0	3/7	4/7
C=1	1/3	2/3

C=0	.7
C=1	.3

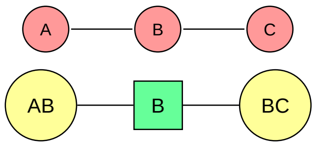
p(A,B,C) = p(A|B)p(B|C)p(C)

Wish to find p(B|A=0) , p(C|A=0)

Problem setup



Transformation of graph



	A=0	A=1
B=0	3/4	1/4
B=1	2/3	1/3

B=0	1
B=1	1

	C=0	C=1
B=0	.3	.1
B=1	.4	.2

A|B	A=0	A=1
B=0	3/4	1/4
B=1	2/3	1/3

B|C	B=0	B=1
C=0	3/7	4/7
C=1	1/3	2/3

C=0	.7
C=1	.3

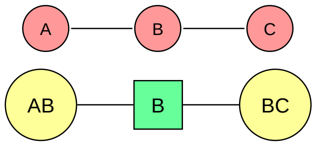
Initialisation of potential representation

We now have a valid potential representation

$$p(X) = \frac{\prod_{cliquesC} \psi(X_C)}{\prod_{separatorsS} \psi(X_S)}$$

$$p(A,B,C) = \frac{\psi(A,B)\psi(B,C)}{\psi(B)}$$

but individual potentials are not yet marginal distributions



	A=0	A=1
B=0	3/4	1/4
B=1	2/3	1/3

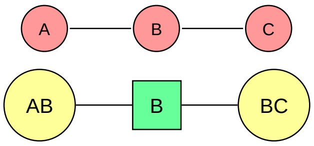
B=0	1
B=1	1

	C=0	C=1
B=0	.3	.1
B=1	.4	.2

	A=0	A=1
B=0	3/4×.4/1	1/4 ×.4/1
B=1	2/3 ×.6/1	1/3 ×.6/1

B=0	.4
B=1	.6

Passing message from BC *to* AB *(1)* multiply marginalise



	A=0	A=1
B=0	.3	.1
B=1	.4	.2

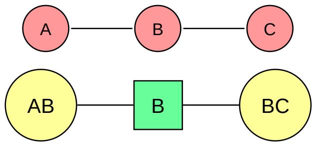
B=0	.4
B=1	.6

	C=0	C=1
B=0	.3	.1
B=1	.4	.2

	A=0	A=1
B=0	3/4×.4/1	1/4 ×.4/1
B=1	2/3 ×.6/1	1/3 ×.6/1

B=0	.4
B=1	.6

Passing message from BC *to* AB *(2)* assign



	A=0	A=1
B=0	.3	.1
B=1	.4	.2

B=0	.4
B=1	.6

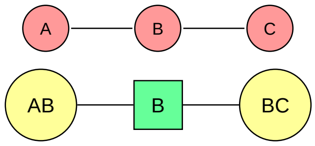
	C=0	C=1
B=0	.3	.1
B=1	.4	.2

After equilibration - marginal tables

We now have a valid potential representation where individual potentials *are* marginals:

$$p(X) = \frac{\prod_{cliques C} p(X_C)}{\prod_{separators S} p(X_S)}$$

$$p(A,B,C) = \frac{p(A,B)p(B,C)}{p(B)}$$



	A=0	A=1
B=0	.3	0
B=1	.4	0

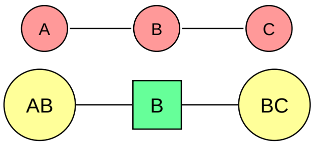
B=0	.4
B=1	.6

	C=0	C=1
B=0	.3	.1
B=1	.4	.2

B=0	.3
B=1	.4

	C=0	C=1
B=0	.3×.3/.4	.1×.3/.4
B=1	.4×.4/.6	.2×.4/.6

Propagating evidence (1)



	A=0	A=1
B=0	.3	0
B=1	.4	0

B=0	.3
B=1	.4

	C=0	C=1
B=0	.225	.075
B=1	.267	.133

B=0	.3
B=1	.4

	C=0	C=1
B=0	.3×.3/.4	.1×.3/.4
B=1	.4×.4/.6	.2×.4/.6

Propagating evidence (2)

We now have a valid potential representation

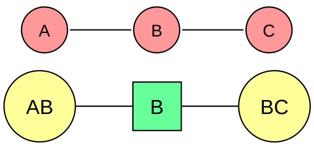
$$p(X) = \frac{\prod_{cliques C} \psi(X_C)}{\prod_{separators S} \psi(X_S)}$$

$$p(A,B,C) = \frac{\psi(A,B)\psi(B,C)}{\psi(B)}$$

where

$$\psi(X_E) = p(X_E \cap \{A=0\})$$

for any clique or separator E



	A=0	A=1
B=0	.3	0
B=1	.4	0

B=0	.3
B=1	.4

	C=0	C=1
B=0	.225	.075
B=1	.267	.133

B=0	B=1	total
.3	.4	.7
.429	.571	

C=0	C=1	total
.492	.208	.7
.702	.298	

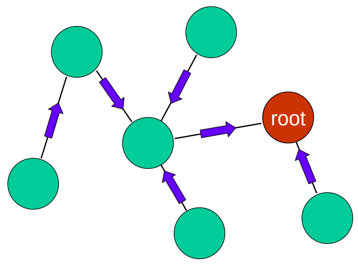
Propagating evidence (3)

Scheduling messages

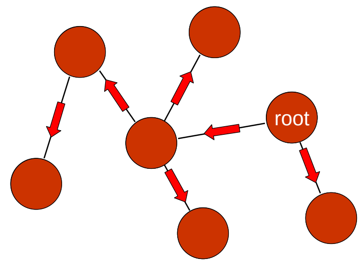
There are many valid schedules for passing messages, to ensure convergence to stability in a prescribed finite number of moves.

The easiest to describe uses an arbitrary root-clique, and first collects information from peripheral branches towards the root, and then distributes messages out again to the periphery

Scheduling messages



Scheduling messages



Scheduling messages

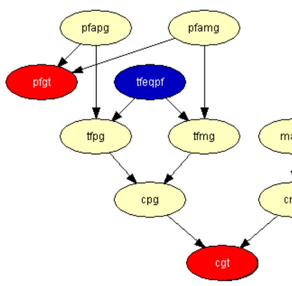
When 'evidence' is introduced - the value set for a particular node, all that is needed to propagate this information through the graph is to pass messages out from that node.

D. Applications

An example from forensic genetics

DNA profiling based on STR's (single tandem repeats) are finding many uses in forensics, for identifying suspects, deciding paternity, etc. Can we use Mendelian genetics and Bayes' theorem to make probabilistic inference in such cases?

Graphical model for a paternity enquiry - neglecting mutation

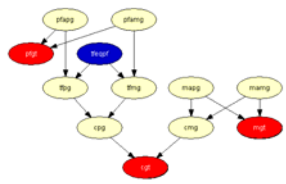


Having observed the genotype of the child, mother and putative father, is the putative father the true father?

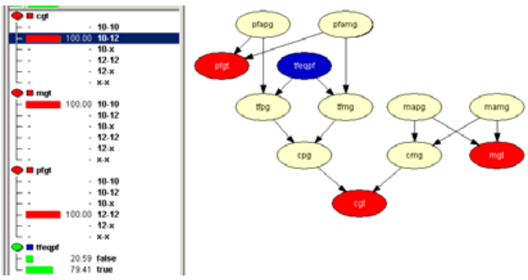
Graphical model for a paternity enquiry - neglecting mutation

Having observed the genotype of the child, mother and putative father, is the putative father the true father?

Suppose we are looking at a gene with only 3 alleles - 10, 12 and 'x', with population frequencies 28.4%, 25.9%, 45.6% - the child is 10-12, the mother 10-10, the putative father 12-12



Graphical model for a paternity enquiry - neglecting mutation



⇒ we're 79.4% sure the putative father is the true father

Surgical rankings

- 12 hospitals carry out different numbers of a certain type of operation:

  47, 148, 119, 810, 211, 196, 148, 215, 207, 97, 256, 360 respectively.
- They are differently successful, and there are:

  0, 18, 8, 46, 8, 13, 9, 31, 14, 8, 29, 24 fatalities, respectively.

Surgical rankings, continued

- What inference can we draw about the relative qualities of the hospitals based on these data?

- A natural model is to say the number of deaths y_i in hospital i has a Binomial distribution $y_i \sim Bin(n_i,p_i)$ where the n_i are the numbers of operations, and it is the p_i that we want to make inference about.

Surgical rankings, continued

- How to model the p_i?
- We do not want to assume they are all the same.
- But they are not necessarily `completely different'.
- In a Bayesian approach, we can say that the p_i are random variables, drawn from a common distribution.

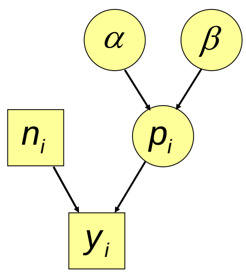
Surgical rankings, continued

- Specifically, we could take

$$\log \frac{p_i}{1-p_i} \sim Beta(\alpha,\beta)$$

- If α and β are fixed numbers, then inference about p_i only depends on y_i (and n_i, α and β).

Graph for surgical rankings

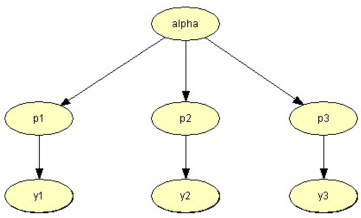


Surgical rankings, continued

- But don't you think that knowing that p_1=0.08, say, would tell you *something* about p_2?
- Putting prior distributions on α and β allows `borrowing strength' between data from different hospitals

Surgical rankings - simplified

3 hospitals, *p* discrete, only one hyperparameter



Surgical rankings - simplified

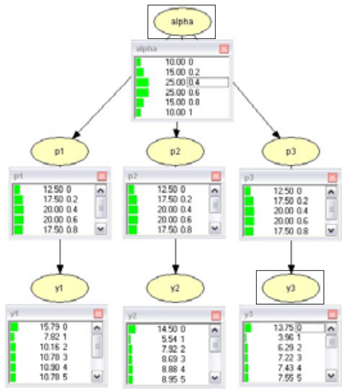
prior for α

alpha	Numbered
0	0.1
0.2	0.15
0.4	0.25
0.6	0.25
0.8	0.15
1	0.1

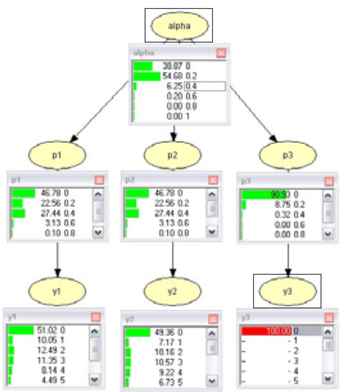
prior for p_i given α

p1 / alpha	0	0.2
0	0.5	0.5
0.2	0.5	0
0.4	0	0.5
0.6	0	0
0.8	0	0
1	0	0

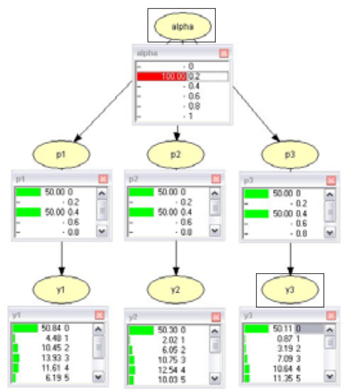
Surgical rankings



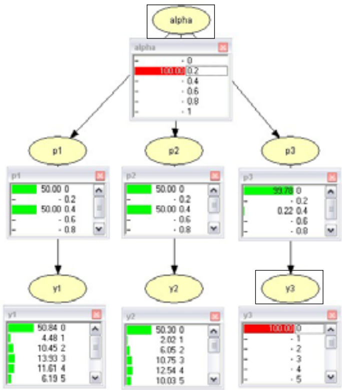
Surgical rankings



Surgical rankings



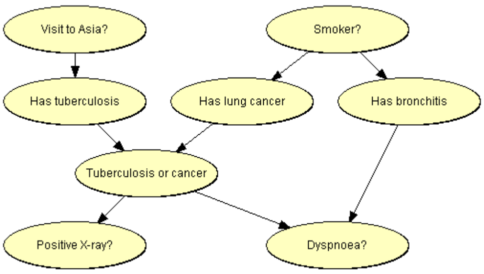
Surgical rankings



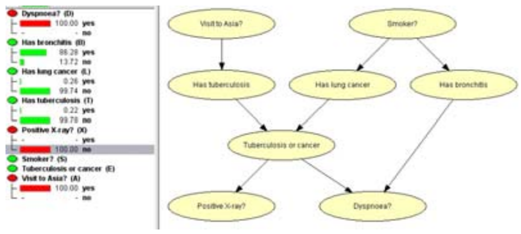
The ‘Asia’ (chest-clinic) example

Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, more than one of these diseases or none of them. A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.

Visual representation of the Asia example - a graphical model



The ‘Asia’ (chest-clinic) example



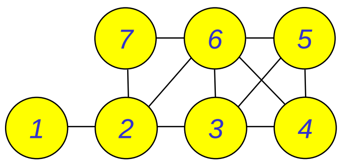
The ‘Asia’ (chest-clinic) example

```
query('asia',c(0.01,0.99))
query('smoke')
tab(c('tb','asia'),,c(.05,.95,.01,.99),c('yes','no'))
tab(c('cancer','smoke'),,c(.1,.9,.01,.99),c('yes','no'))
tab(c('bronc','smoke'),,c(.6,.4,.3,.7),c('yes','no'))
or('tbcanc','tb','cancer')
tab(c('xray','tbcanc'),,c(.98,.02,.05,.95),c('yes','no'))
tab(c('dysp','tbcanc','bronc'),,c(.9,.1,.8,.2,.7,.3,.1,.9),c('yes','no'))

prop.evid('asia','yes')
prop.evid('dysp','yes')
prop.evid('xray','no')
pnmarg('cancer')

 cancer=yes cancer=no
0.002550419 0.9974496
```

Software



- The HUGIN system: freeware version (Hugin Lite 5.7): http://www.stats.bris.ac.uk/~peter/Hugin57.zip
- Grappa (suite of R functions) http://www.stats.bris.ac.uk/~peter/Grappa