PROBABILISTIC REASONING

Unit # 4

MODELING BAYESIAN NETWORKS

KNOWLEDGE ELICITATION

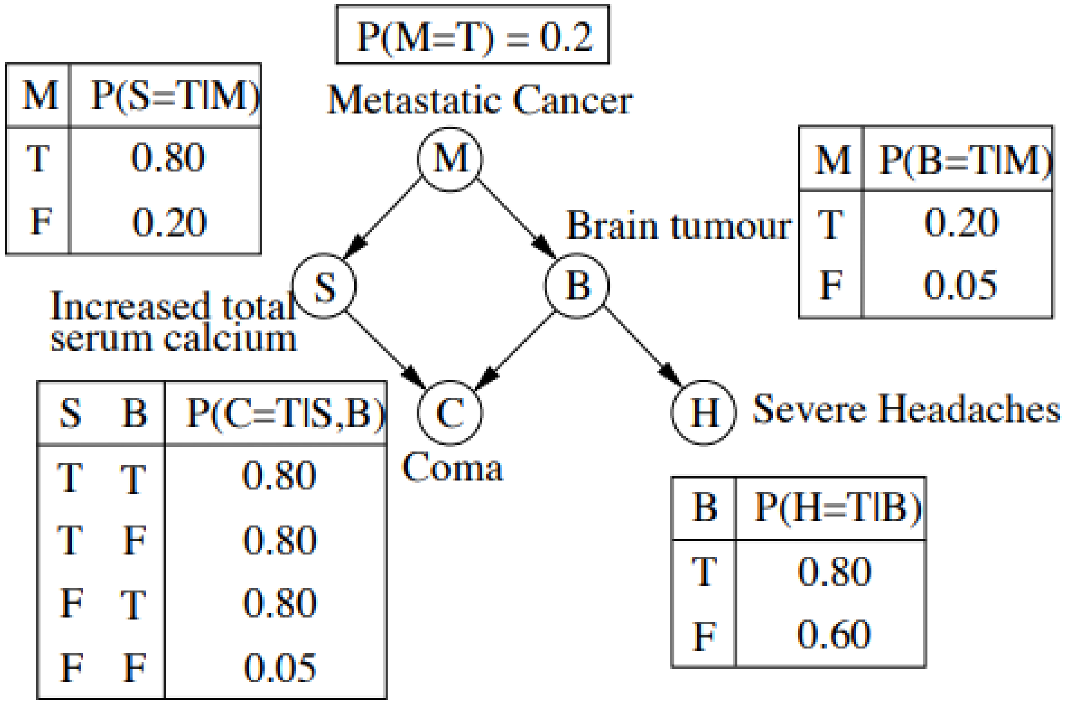
- Types of nodes
  - Target/query nodes
  - Evidence/Observation
- Types of values
  - Boolean
  - Categorical

COMMON MODELING MISTAKES

- Discrete variable values must be ‘exhaustive’ and ‘exclusive’.
- Creation of separate variables for different states of the same node

EXAMPLE STATEMENT: METASTATIC CANCER

Metastatic cancer is a possible cause of brain tumors and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also associated with brain tumors. (This example has a long history in the literature, namely Cooper, 1984, Pearl, 1988, Spiegelhalter, 1986.)



M	P(S=T|M)
T	0.80
F	0.20

M	P(B=T|M)
T	0.20
F	0.05

S	B	P(C=T|S,B)
T	T	0.80
T	F	0.80
F	T	0.80
F	F	0.05

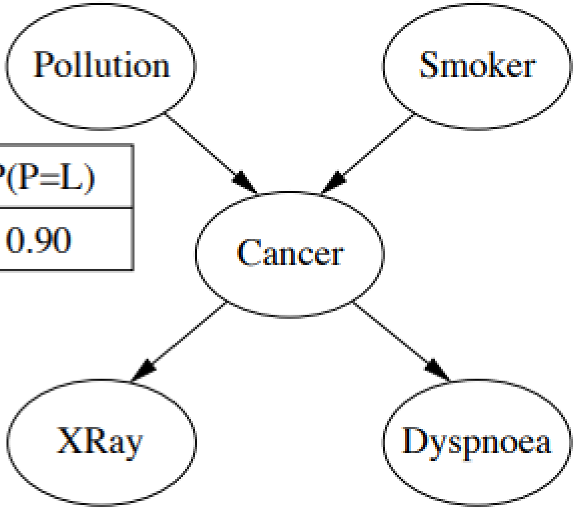
B	P(H=T|B)
T	0.80
F	0.60

FIGURE 2.7: Metastatic cancer BN.

BUILDING A NETWORK IN GENIE

EXAMPLE PROBLEM: LUNG CANCER

A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer



P(S=T)
0.30

P(P=L)
0.90

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	P(X=pos|C)
T	0.90
F	0.20

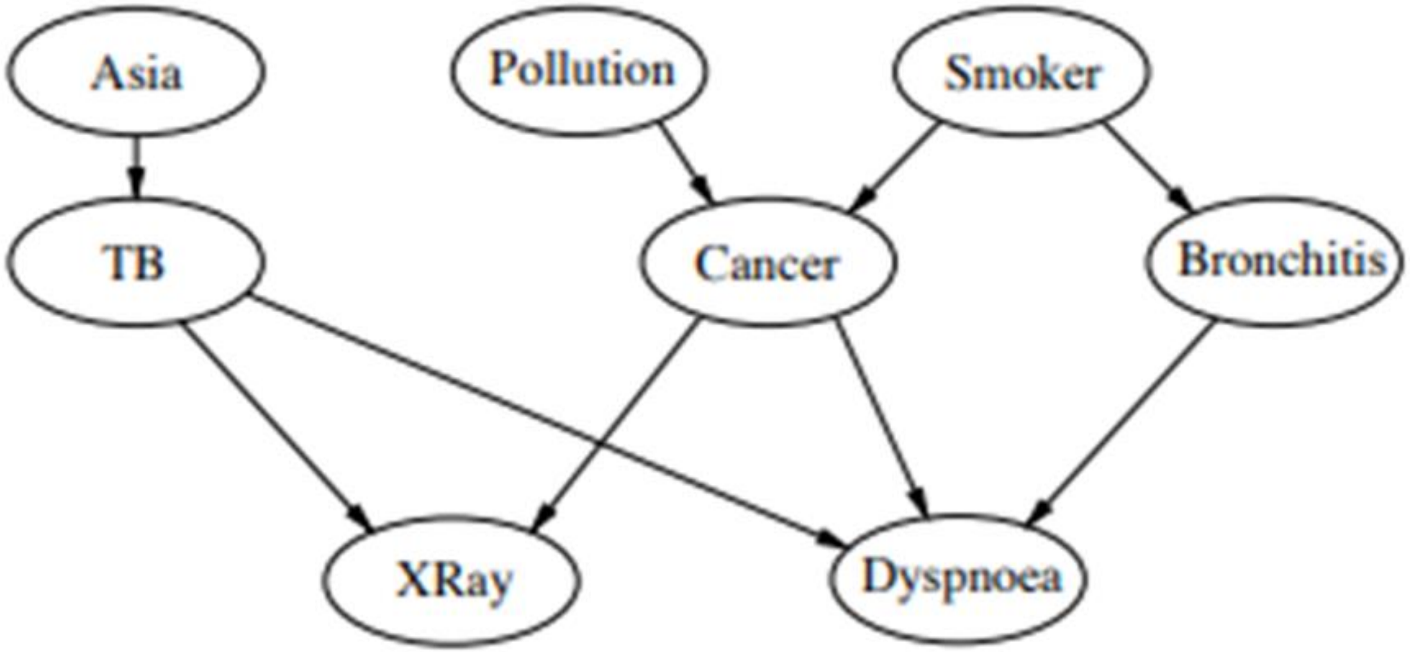
C	P(D=T|C)
T	0.65
F	0.30

FIGURE 2.1: A BN for the lung cancer problem.

EXAMPLE

Suppose that we wanted to expand our original medical diagnosis example to represent explicitly some other possible causes of shortness of breath, namely tuberculosis and bronchitis. Suppose also that whether the patient has recently visited Asia is also relevant, since TB is more prevalent there

MODELING BAYESIAN NETWORKS



ALTERNATIVE FORM

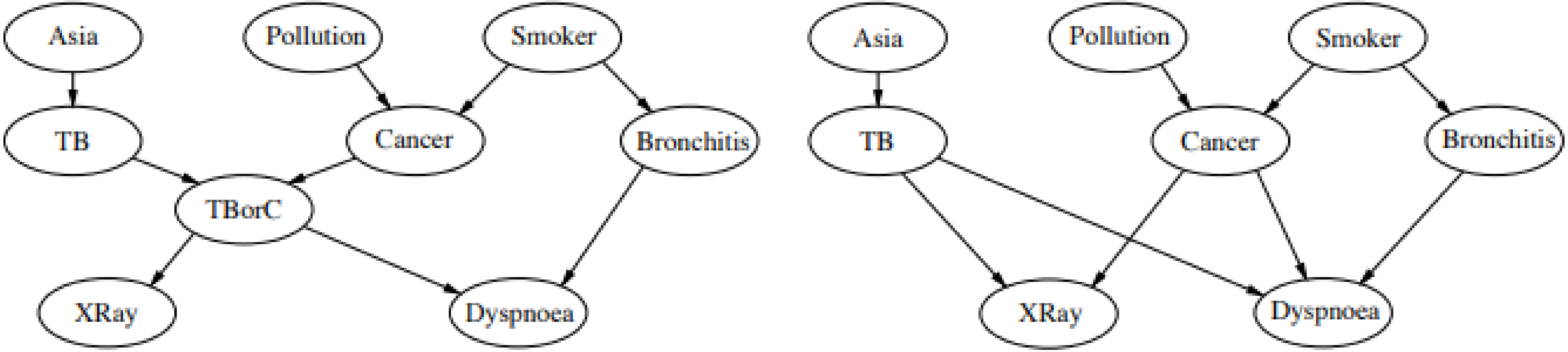


FIGURE 2.8: Alternative BNs for the “Asia” example.

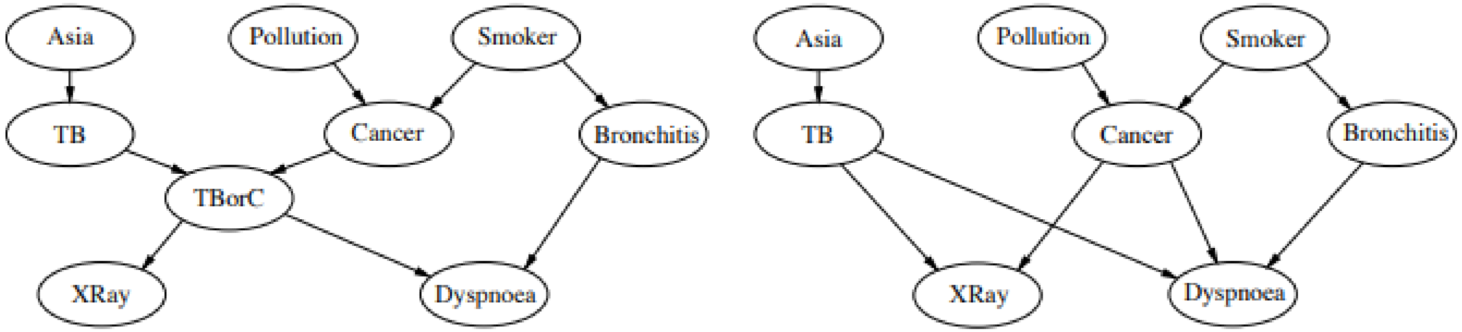
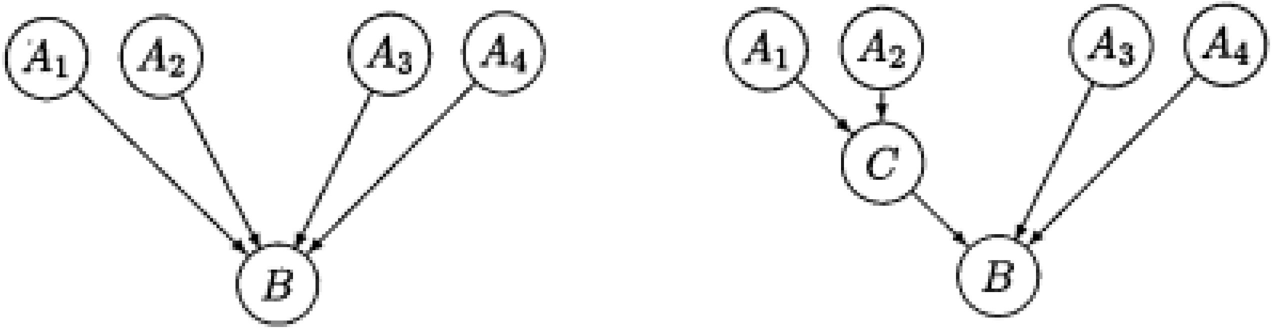


FIGURE 2.8: Alternative BNs for the "Asia" example.

In both networks all the nodes are Boolean. Note the slightly odd intermediate node TBorC, indicating that the patient has either tuberculosis or bronchitis. This node is not strictly necessary; however, it reduces the number of arcs elsewhere, by summarizing the similarities between TB and lung cancer in terms of their relationship to positive X-ray results and dyspnoea. Without this node, as can be seen on the right, there are two parents for X-ray and three for Dyspnoea, with the same probabilities repeated in different parts of the CPT. The use of such an intermediate node is an example of "divorcing," a model structuring method

DIVORCING

To reduce the number of combinations of states for a variable with multiple parents, and hence the number of probabilities to be calculated, the divorcing technique may be employed. Some of the parents of a variable are removed or divorced from that variable by introducing a mediating variable and making it a child of the divorced parents and a parent of the original child. For example, in figure 6.3, the variable C is introduced to divorce the parents A\ and A2 of B. For divorcing to be effective, the number of states of C must be less than the number of combinations of states of Ai and A2.



(a) Original (b) After divorcing

A Tutorial on Bayesian Belief Networks (dtic.mil)

DRIVING EXAMPLE IN GENIE

DIVORCING EXAMPLE (GRANTING A LOAN)

A bank will decide on a mortgage loan for a customer who wishes to purchase a house. The customer is asked to fill in a form giving information on various financial and personal matters together with various key information on the house. The answers are used to estimate the probability that the bank will get its money back. The information can be the following:

- type of job, yearly income, other financial commitments, number and types of cars in the family, number of previous addresses during the last five years, number of children in the family, number of divorces, size and age of the house, price of the house, and type of environment.

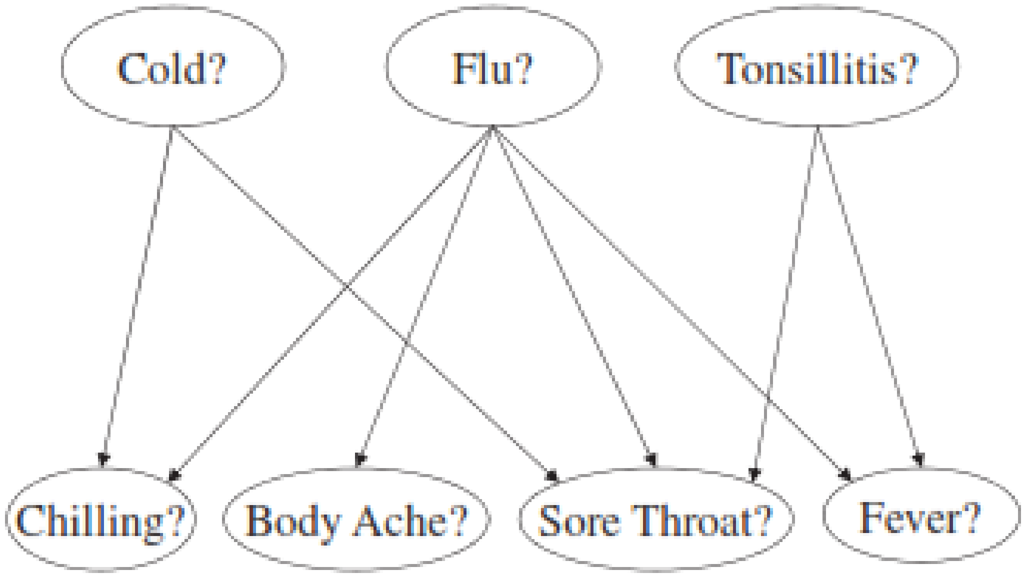
If we assume that each parent variable has five states, we have already listed a parent space with 5^{11} ~ 5,000,000 configurations. For each configuration, we request a distribution for:

- No person can estimate that number of distributions, nor can he or she estimate a distribution for a divorced businesswoman with a yearly income of $50,000, having loans of $70,000 already, one car, three previous addresses, two children, wanting to purchase a twenty-year-old house of 150 m2 at the price of $200,000 in a farming area.
- Also, if the distributions are to be taken from a database, the bank will need at least 50,000,000 cases that may not be more than 10 years old.

EXERCISE

The flu is an acute disease characterized by fever, body aches, and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils that leads to a sore throat and can be associated with fever.

EXAMPLE



CANONICAL MODELS

Canonical models represent the relations between a set of random variables for particular interactions using few parameters.

It can be applied when the probabilities of a random variable in a BN conform to certain *canonical* relations with respect to the configurations of its parents.

There are several classes of canonical models, the most common are the *Noisy OR* and *Noisy AND* for binary variables, and their extensions for multivalued variables, *Noisy Max* and *Noisy Min*, respectively.

NOISY-OR

The Noisy OR is basically an extension of the OR relation in logic.

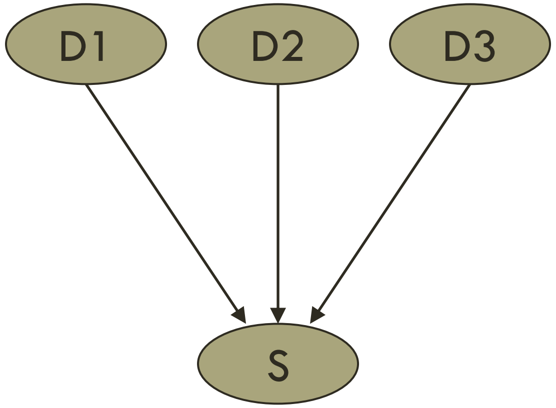
Consider an OR logic gate, in which the output is *True* if any of its inputs are *True*.

The Noisy OR model is based on the concept of the logic OR; the difference is that there is a certain (small) probability that the variable is not *True* even if one or more of its parents are *True*.

The Noisy OR model is applied when several variables or *causes* can produce an *effect* if any one of them is *True*, and as more of the *causes* are true, the probability of the effect increases.

EXAMPLE

For instance, the effect could be a certain symptom, *S*, and the causes are a number of possible diseases, *D*1, *D*2, . . . , *Dm*, that can produce the symptom, such that if none of the diseases is present (all *False*) the symptom does not appear; and when any disease is present (*True*) the symptom is present with high probability and increases as the number of *Di* = *True* increases.



CONDITIONS

The following two conditions must be satisfied for a NoisyOR canonical model to be applicable:

Responsibility: the effect is false if all the possible causes are false.

Independence of exceptions: if an effect is the manifestation of several causes, the mechanisms that inhibit the occurrence of the effect under one cause are independent of the mechanisms that inhibit it under the other causes.

NOISY-OR EQUATIONS

If $A_1, A_2, \ldots\ldots\ldots, A_n$ are the causes of a variable B, then the probability of each A_i causing B is p_i.

The probability of Ai inhibiting B is q_i or $(1- p_i)$.

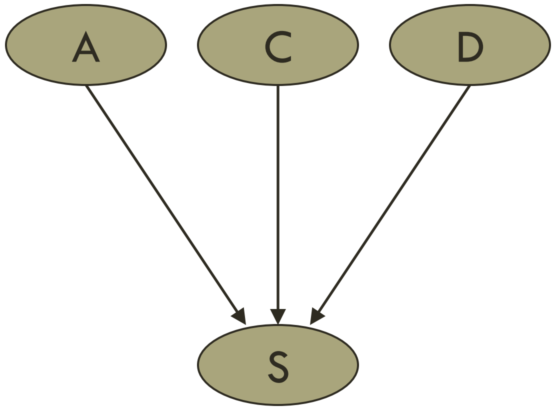
Assuming that the inhibiting variables are independent, the combined probability becomes one minus the product of the probability of inhibitors.

$$P\,(B \mid A_1, A_2, \ldots\ldots\ldots, A_n) = 1 - (1- p_1)(1- p_2)\ldots\ldots(1- p_n)$$

EXAMPLE (LASKEY)

Sneezing can be caused by an allergy (A), a cold (C), or dust (D) in the air

- Allergy triggers sneezing with probability pA = .6
- Cold triggers sneezing with probability pC = .9
- Dust triggers sneezing with probability pD = .3



NOISY OR - INHIBITIONS

Inhibition in the Noisy OR model is important because it addresses the complexity of how multiple causes interact to affect the probability of an outcome.

By modeling inhibition, the Noisy OR model can more accurately represent real-world situations where causes are not independent and their effects can overlap or interfere with each other.

This allows the model to provide a more nuanced and realistic understanding of the relationships between causes and outcomes.

NOISY-OR COMPUTATION

P(Sneezing | Allergy) = 0.6

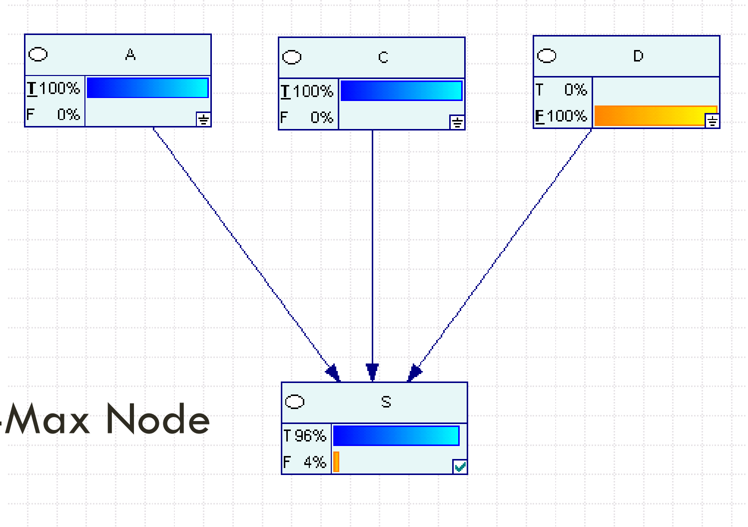
P(Sneezing | Cold) = 0.9

P(Sneezing | Dust) = 0.3

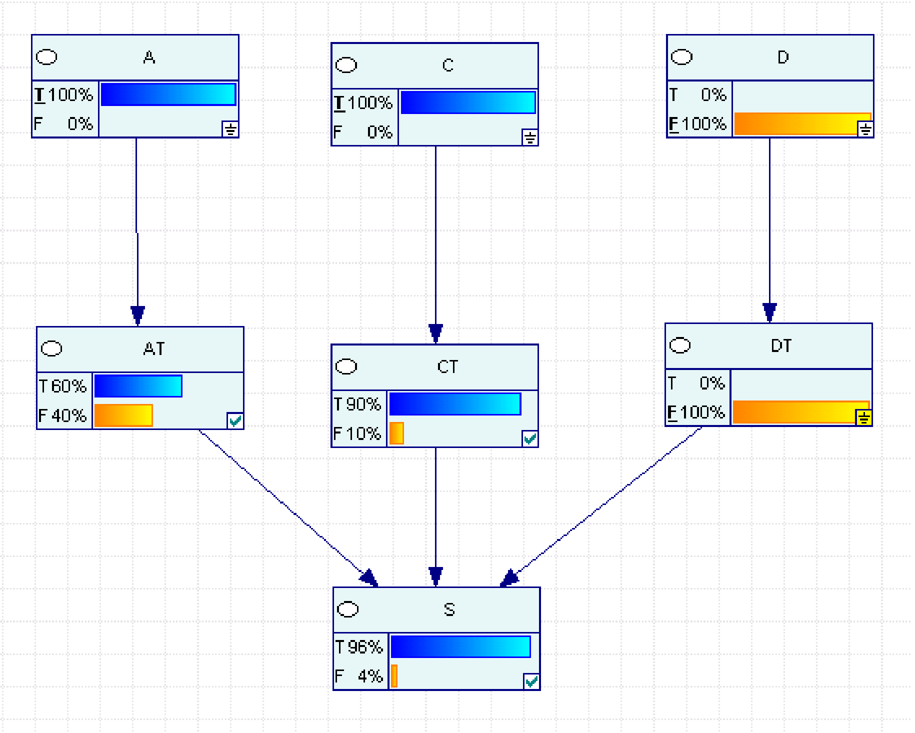
A	C	D	P(S)
F	F	F	
F	F	T	
F	T	F	
F	T	T	
T	F	F	
T	F	T	
T	T	F	
T	T	T	

NOISY-OR MODELING USING GENIE

NOISY OR: ALTERNATE REPRESENTATION



Noisy-Max Node



LEAK PROBABILITY

Noisy-OR model can be extended to incorporate the probability that the effect can be produced by some un-modeled cause.

This requires the elicitation of one more probability, called "leak" probability, from the expert.

To include the leak probability, the equation can be modified as

$$P\ (B \mid A_1, A_2, \ldots\ldots\ldots, A_n\ ) = 1 - (1 - p_0)(1 - p_1)(1 - p_2)\ldots\ldots.(1 - p_n)$$

where p0 is the probability that B will be true even if none of the A_i are true.

NOISY-OR WITH LEAK PROBABILITY

P(Sneezing | Allergy) = 0.6

P(Sneezing | Cold) = 0.9

P(Sneezing | Dust) = 0.3

If none of "known" causes are present, a person could still have sneezing.

$P(S \mid \neg A, \neg C, \neg D) = 0.01$

A	C	D	P(S)
F	F	F	
F	F	T	
F	T	F	
F	T	T	
T	F	F	
T	F	T	
T	T	F	
T	T	T	