Harvard-MIT Division of Health Sciences and Technology
HST.950J: Engineering Biomedical Information: From Bioinformatics to Biosurveillance
Course Directors: Dr. Isaac Kohane, Dr. Marco Ramoni

Harvard Medical School

Massachusetts Institute of Technology

Biomedical Computing

Decision Support Systems
November 3rd, 2005

6.872 / HST 950

Decision Support Systems

- An intelligent system is a computer program able to emulate intelligent performances.
- Computer program:
  - ✓ A mechanical (effective) procedure.
- Intelligent performances (pragmatic definition):
  - ✓ The performance we expect to require intelligence.
- Emulate/Simulate:
  - ✓ Emulate means to achieve the same objectives.
  - ✓ Simulate means to reproduce the same behavior.

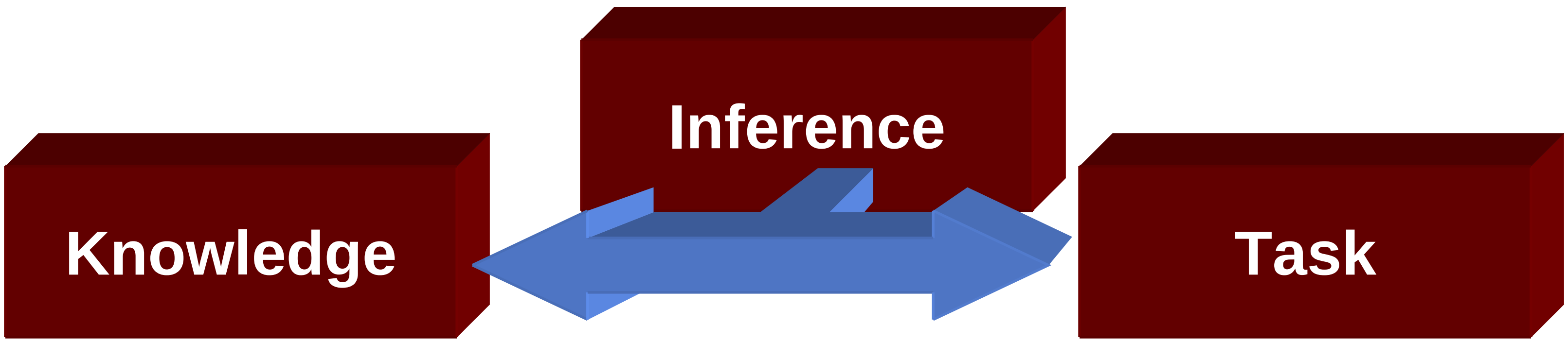
Components

Knowledge: Representation of domain knowledge.

- You may regard this knowledge as axioms.

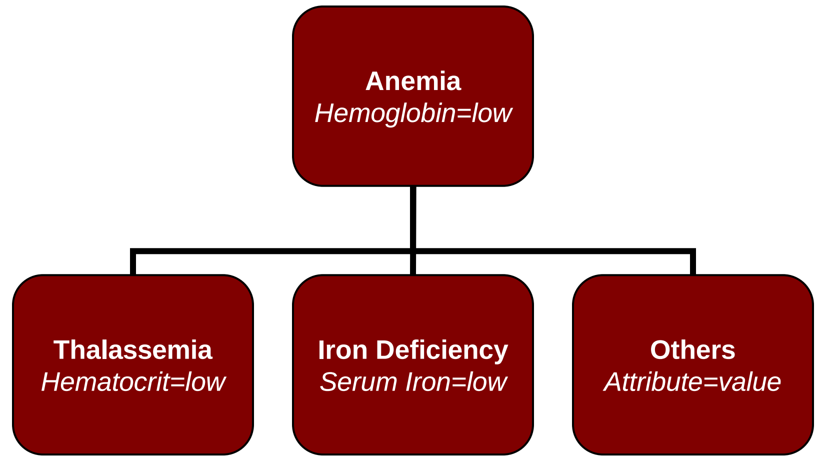
Inference: Domain independent procedures to handle knowledge in order to achieve these tasks.

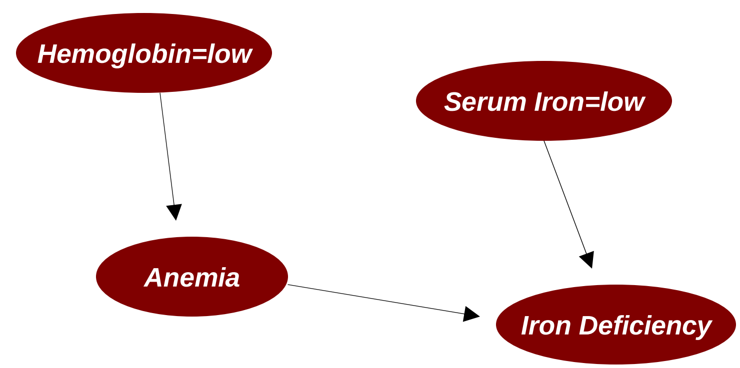
- You may regard these procedures as inference rules.



Knowledge Representation

- Knowledge/Information
  - ✓ Knowledge is not just information.
  - ✓ Knowledge is structured information.
- Ontology is the structure of the domain knowledge.
  - ✓ A classification of medical disorders: a hierarchy prototype definitions.
  - ✓ A network of causal relationships and influences.





Inference and Reasoning

- Once knowledge is represented, it must be used.
- Inference: The operation able to draw conclusions.
  - Sound: draw only true conclusions.
  - Complete: draw all the true conclusions.
  - *The truth, the whole truth, and nothing but the truth*.
- Reasoning: Application of inference to knowledge.
  - Truth preservative: draw true from true.
  - Monotonic: the conclusions drawn are always valid (!).
  - *The first 33 theorems of Euclid's Elements are drawn without the Fifth Axiom but they still hold after it*.

Problem Solving

Knowledge/inference compilation:

- if (the infection is meningitis) 1
  - (the type is bacterial) 2
  - (therapy is corticosteroids) 3
  - (only circumstatial evidence) 4

  then
  - klebsiella (0.2), e.coli (0.4) 5
  - or pseudomonas (0.1) 6

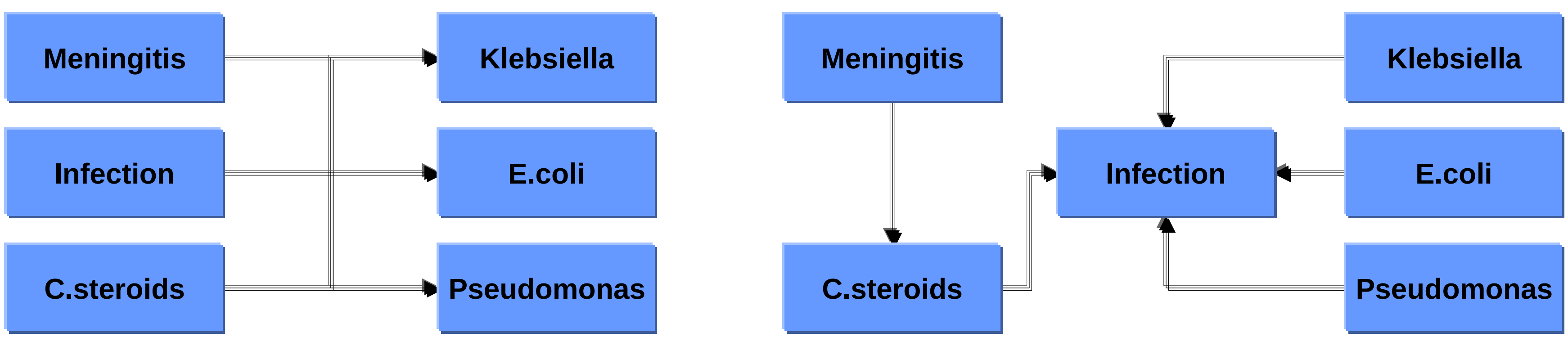
- ✓ Knowledge (1 2 3 5 6) and inference (2).
- ✓ Difficult to acquire, maintain, and update.

Deep systems:

- ✓ Knowledge flows in the opposite direction.
- ✓ Inference must reverse this natural path of knowledge.

Logical Representation

- Logical knowledge representation (axioms):
  - (meningitis ⇒ corticosteroids)
    (corticosteroids ⇒ bacterial_infection)
    (klebsiella ⇒ bacterial_infection)
    (e.coli ⇒ bacteria_infection)
    (pseudomonas ⇒ bacterial_infection)
  - (circumstantial evidence)
    (0.2) (0.4) (0.1)



Logical Reasoning

- Inference rules (e.g. modus [ponendi] ponens)
  - Metalinguistic inference rules:
    - $\alpha \Rightarrow \beta$
      α
      $\therefore \beta$
  - Axiom schema:
    - $(\alpha \Rightarrow \beta) \wedge \alpha \Rightarrow \beta$
  - Example:
    - Axioms:
      - (cat ⇒ mammal); cat
    - Application of the inference rule:
      - cat ⇒ mammal
        cat
        ∴ mammal

Example

- Axioms
  - (meningitis ⇒ corticosteroids)
    (corticosteroids ⇒ bacterial_infection)
    (klebsiella ⇒ bacterial_infection)
    (e.coli ⇒ bacterial_infection)
    (pseudomonas ⇒ bacterial_infection)
- Observations:
  - meningitis
- Inference
  - ☞meningitis ⇨ corticosteroids ⇨ bacterial_infection

☹ But how do we infer the type of bacteria?

Limits

- Representation of uncertainty:
  - How do we represent uncertainty about knowledge?
    - Meningitis may cause brain damage.
- Abductive inference:
  - How do we get the bacteria?
    - cat $\Rightarrow$ mammal
      mammal
      $\therefore$ cat
    - e.coli => bacteria
      bacteria
      $\therefore$ e.coli
- *This pattern is called Fallacy of Affirming the Consequent*.

Probability

- Let's consider our propositions as events:
  - ✓ Probability is a function mapping an event to [0 1].

$$0 \leq p(a) \leq 1$$

☞ The probability that tomorrow will rain is 0.4: $p(\text{rain})=0.4$.

- Properties:
  - ✓ Complementation:

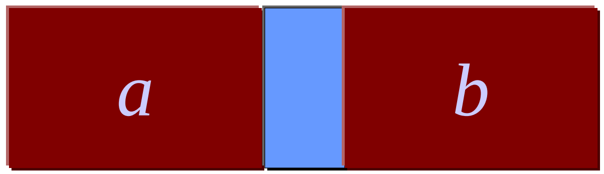
$$p(a) + p(\neg a) = 1$$

$$p(\neg a) = 1 - p(a)$$

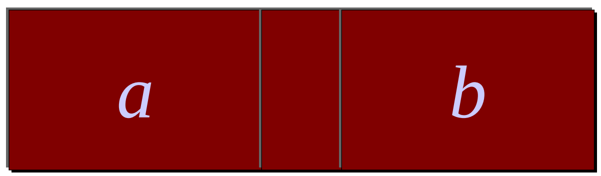
☞ The probability that tomorrow will *not* rain: $p(\neg\text{rain})=0.6$.

And and Or

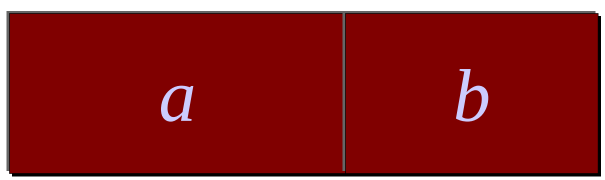
- And ($\wedge$) is set intersection: $p(a \wedge b)$



- Or ($\vee$) is set union: $p(a \vee b)$



- If the intersection is empty, the union is the sum: $p(a \wedge b)=0 \Rightarrow p(a \vee b)=p(a)+p(b)$



Exhaustivity and Exclusivity

Exclusivity: *a* and *b* cannot be both true.

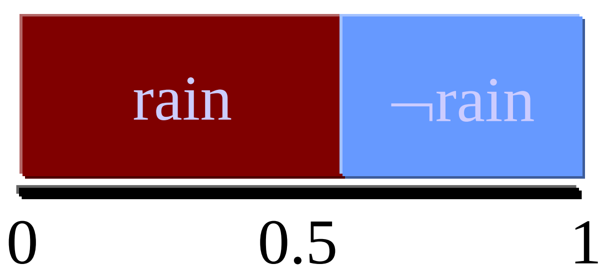
$$p(a \wedge b) = 0$$

When it rains, it may never be the case it is not raining.

Exhaustivity: Events exhaust all the possibilities:

$$p(a) + p(b) = 1$$

Either it is raining or it is not raining.



Axioms

Axiom 1: $p(a) \geq 0$

- ✓ The probability of an event cannot be negative.

Axiom 2: $p(\Omega)=1$

- ✓ The probability of an exhaustive set is 1.
  - ☞ $p(\text{rain})+p(\neg\text{rain})=1$
- ✓ This, together with Axiom 1, implies $0 \leq p(a) \leq 1$.

Axiom 3: $p(a \vee b)= p(a)+p(b)$ if $p(a \wedge b)=0$.

- ✓ Theorem: $p(a \vee b)= p(a)+p(b) - p(a \wedge b)$.

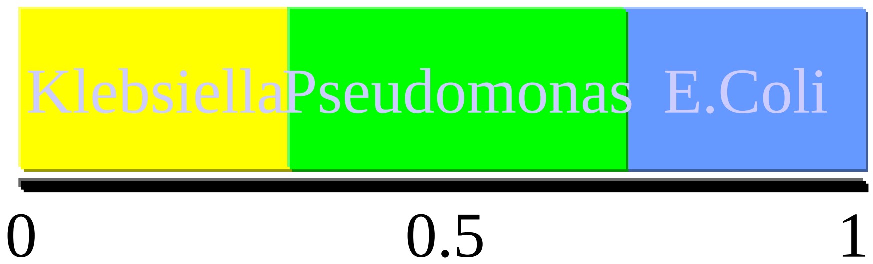
Variables

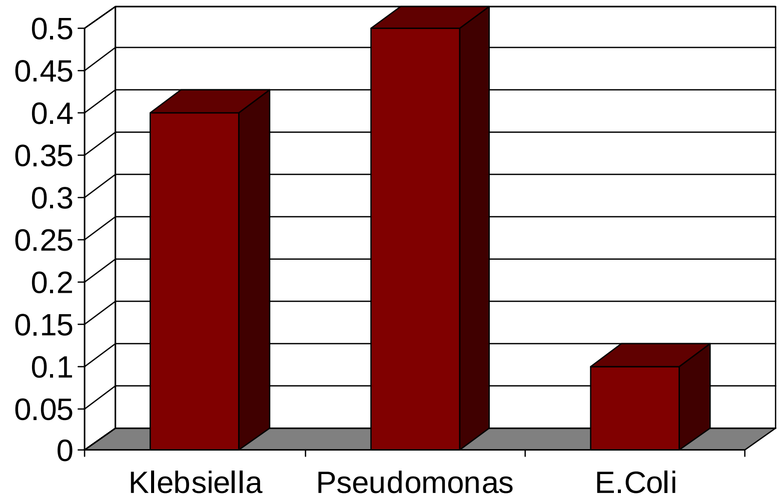
A variable is a symbol *A* with a set of *J* possible values $\{a_1, \ldots, a_J\}$.

- ✓ A state a_{ij} is the assignment of a value to the variable.
- ✓ A variable can take just one value at the time.
- ✓ A variable must take at least one value.
- ✓ A variable is a set of exclusive and exhaustive states.
  - ☞ The variable Bacteria has values: {klebsiela, e.coli, psedumonas}.
  - ☞ There are three states, such as Bacteria= klebsiela.
- ✓ Our propositions are variables with just two states.
  - ☞ The variable Rain can take two values: {true,false}.
  - ☞ Rain has two states: Rain=true and Rain=false.

Probability Distribution

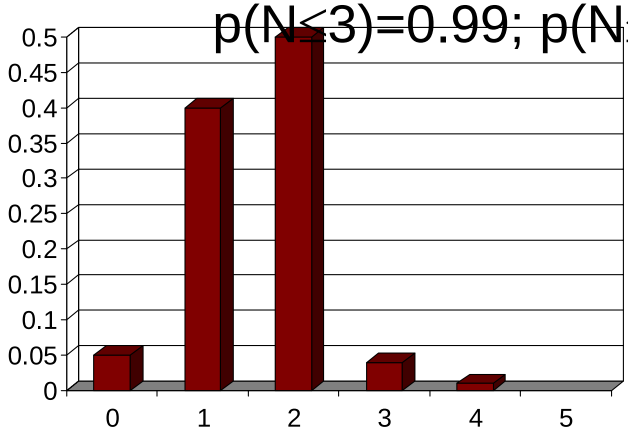
- We can associate a probability value to each state.
- The set of probabilities associated to a variable is called Probability Distribution.
  - ✓ Since the states are exhaustive, they sum up to 1.
  - ✓ Since the states are exclusive, their intersections are 0.

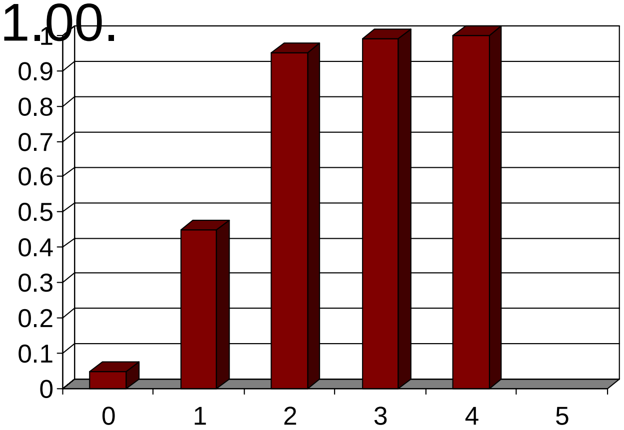




Cumulative Distributions

- The cumulative distribution is given by the sum of probabilities of some states.
  - The variable N *Number of assignments per course.*
    - p(N=0)=0.05; p(N=1)=0.40; p(N=2)=0.50; p(N=3)=0.04; p(N=4)=0.01; p(N≥5)=0.00.
  - The Cumulative Distribution of N.
    - p(N≤0)=0.05; p(N≤1)=0.45; p(N≤2)=0.95; p(N≤3)=0.99; p(N≤4)=1.00.





Joint Probability

Conjunction is represented as joint probability.

- $p(\text{cloudy} \wedge \text{rain}) = 0.60$ 1
  $p(\text{cloudy} \wedge \neg\text{rain}) = 0.15$ 2
  $p(\neg\text{cloudy} \wedge \text{rain}) = 0.15$ 3
  $p(\neg\text{cloudy} \wedge \neg\text{rain}) = 0.10$ 4

Imagine joint events as percentages of days in a year. Note that they sum up to 1.

Inference (marginal probability):

- $p(\text{cloudy}) = 1 + 2 = 0.60 + 0.15 = 0.75$
  $p(\text{rain}) = 1 + 3 = 0.60 + 0.15 = 0.75$
  $p(\neg\text{cloudy}) = 3 + 4 = 0.15 + 0.10 = 0.25$
  $p(\neg\text{rain}) = 2 + 4 = 0.15 + 0.10 = 0.25$

Conditioning

- Modus ponens needs a conditional statement.
- Inference needs conditional statements but we cannot use standard implication ($\Rightarrow$):

  $p(a \Rightarrow b) = 0.7 \equiv p(\neg a \vee b) = 0.7 \equiv p(b) \leq 0.7$ and $p(a) \leq 0.3$ 💣

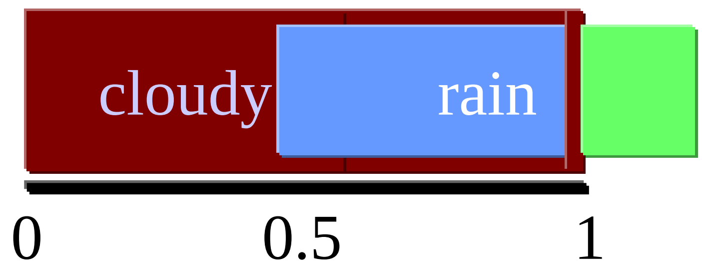
- We need a probabilistic conditioning.
- Fortunately, we have conditional probability:

$$p(a|b) = \frac{p(a \wedge b)}{p(b)}$$

- It is read *the probability of a given b.*

Conditional Probability

- Conditional probabilities represent the intension:
  - ☞If it is cloudy it is more likely to rain
    - ➢ p(rain|cloudy)=0.8; p(¬rain|cloudy)=0.2
  - ☞If it is not cloudy it is more likely not to rain.
    - ➢ p(rain|¬cloudy)=0.6; p(¬rain|¬cloudy)=0.4
- The conditional probability is the probability of an event when another event is happening.
- It is a set zoomed in another set.



Bayes' Theorem

Bayes' Theorem is the solution to abduction:

$$p(a \mid b) = \frac{p(a \wedge b)}{p(b)}$$

- ✓ Since $p(a \wedge b) = p(a \mid b) \times p(b) = p(b \mid a) \times p(a)$:

$$p(b \mid a) = \frac{p(a \mid b)\, p(b)}{p(a \wedge b) + p(a \wedge \neg b)}$$

- ✓ Invert $p(a \mid b)$ into the *posterior probability* $p(b \mid a)$.

$$p(b \mid a) = \frac{p(a \mid b)\, p(b)}{p(a \mid b)\, p(b) + p(a \mid \neg b)\, p(\neg b)}$$

Exhaustive

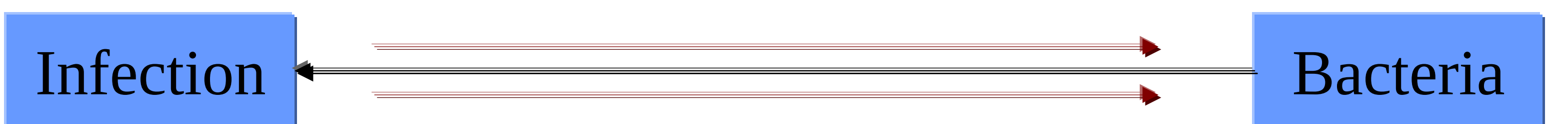
- The bacteria are values of a single variable **B**:
  - **p(B=k)=0.4;p(B=e)=0.3;p(B=p)=0.3**
  - **p(I=t|B=k)=.9;p(I=t|B=e)=.8;p(I=t|B=p)=.8**
    p(I=f|B=k)=.1;p(I=f|B=e)=.2;p(I=f|B=p)=.2
- Posterior probability *p(B=k|I=t)*

$$\frac{p(I=t|B=k)\times p(B=k)}{p(I=t|B=k)\times p(B=k)+p(I=t|B=e)\times p(B=e)+p(I=t|B=p)\times p(B=p)}=$$

$$=\frac{0.9\times 0.4}{0.9\times 0.4+0.8\times 0.3+0.8\times 0.3}=\frac{0.36}{0.84}=0.428$$

$$p(B=e|I=t)=\frac{0.24}{0.84}=0.286 \qquad p(B=p|I=t)=\frac{0.24}{0.84}=0.286$$

$$p(B=k|I=t)+p(B=e|I=t)+p(B=p|I=t)=0.428+0.286+0.286=1$$

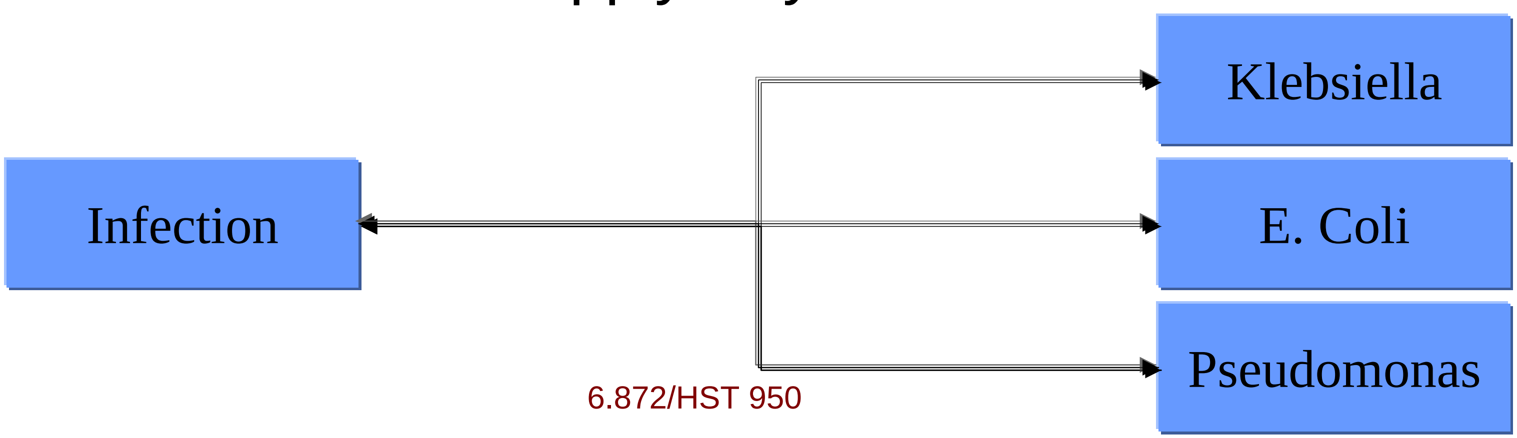


Independent

The bacteria are independent binary variables:

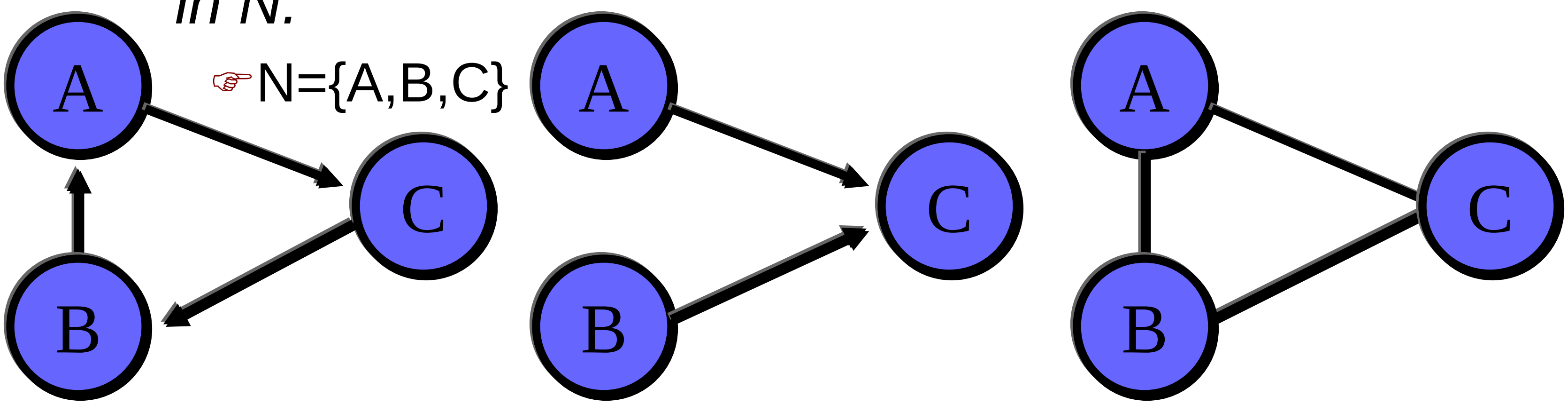
- $p(K=t)=0.4; p(E=t)=0.3; p(P=t)=0.3$
- $p(I=t \mid K=t \wedge E=t \wedge P=t)=1; p(I=t \mid K=t \wedge E=t \wedge P=f)=.9;$
  $p(I=t \mid K=t \wedge E=f \wedge P=t)=.9; p(I=t \mid K=f \wedge E=t \wedge P=t)=.8;$
  $p(I=t \mid K=t \wedge E=f \wedge P=f)=.8; p(I=t \mid K=f \wedge E=t \wedge P=f)=.7;$
  $p(I=t \mid K=f \wedge E=f \wedge P=t)=.7; p(I=t \mid K=f \wedge E=f \wedge P=f)=.1.$

Solution: Transform the antecedents in a single variable: antecedents defines a set of mutually exclusive and exhaustive states and Apply Bayes' Theorem.



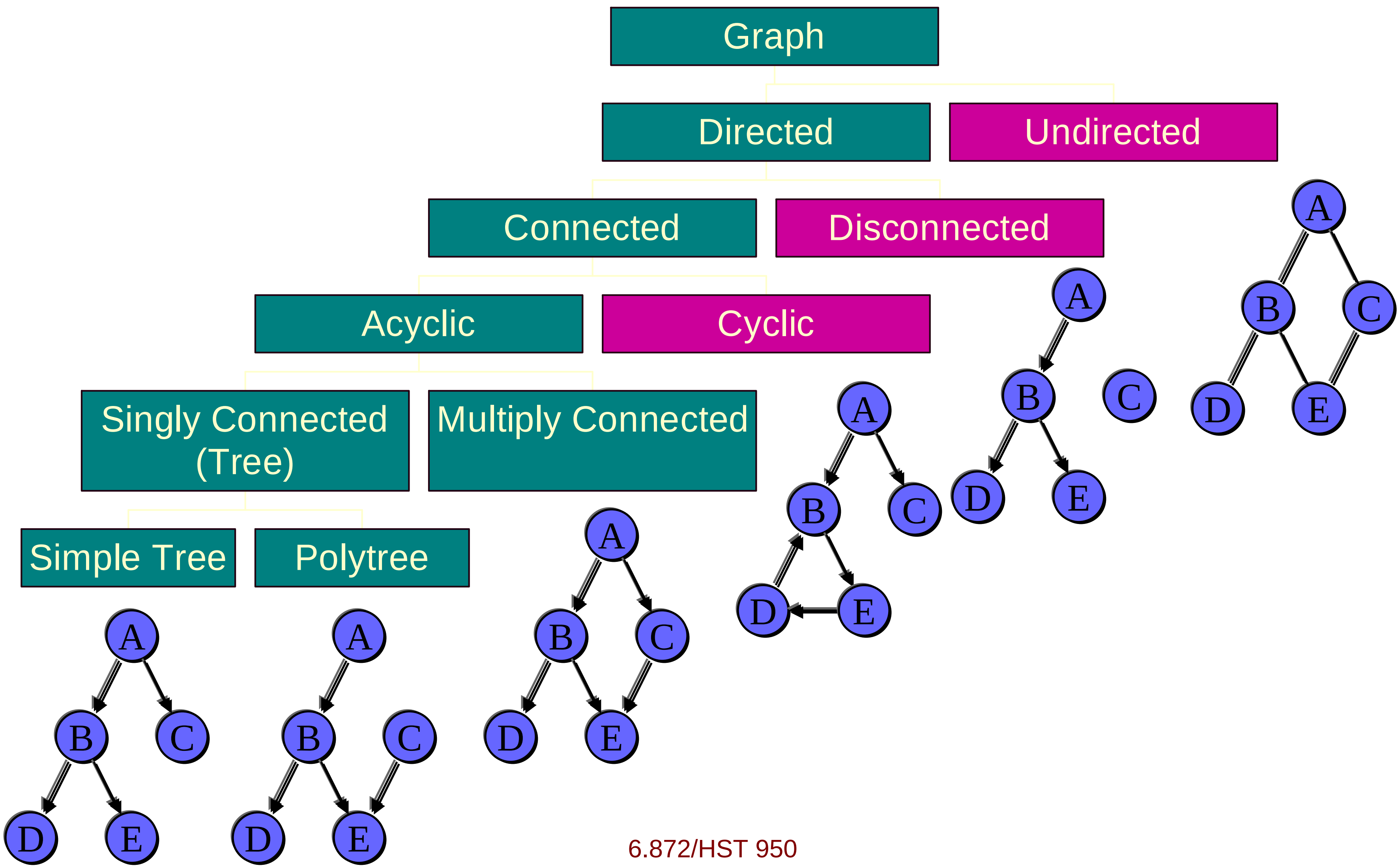
Graph

- A graph (network) *G*(*N*,*L*) is defined by:
  - ✓ A finite set *N* = {*A*,*B*,...} of nodes (vertices).
  - ✓ A set *L* of links (edges): ordered pair of nodes (*A*,*B*).
  - *The set L is a subset of all possible pairs of nodes in N.*
    - N={A,B,C}



L={(A,C),(B,C),(B,A)} L={(A,C),(B,C)} L={(A,C),(B,C),(B,A),(C,A),(C,B),(A,B)}

Types of Graph



Independence

- Two variables are independent ($I(A,B)$) if knowing one does not affect our belief in the other:

$$p(A|B)=p(A)$$

- The conjunction is exactly the Boolean conjunction, since

$$p(A \wedge B)/\mathrm{p(B)}=p(A)$$

therefore:

$$p(A \wedge B)=p(A) \times p(B)$$

  - ☞ p(Rain)=0.2   p(Sprinkler)=0.9
    p(Rain∧Sprinkler)=0.18
    p(Rain)=0.2   p(¬Sprinkler)=0.1
    p(Rain∧¬Sprinkler)=0.02
    p(¬Rain)=0.8 p(Sprinkler)=0.9
    p(¬Rain∧Sprinkler)=0.72
    p(¬Rain)=0.8 p(¬Sprinkler)=0.1
    p(¬Rain∧¬Sprinkler)=0.08

Conditional Independence

- When two variables are independent given a third, they are said to be conditionally independent.

$$p(A|B \wedge C)=p(A \wedge B \wedge C)/p(B \wedge C)=p(A|C)$$

T-shirt size of kids affect their literacy skills.

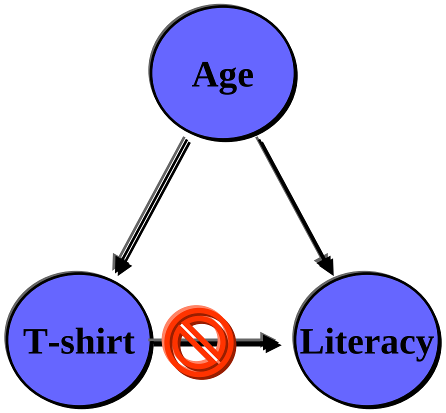
Both T-shirt size and literacy skills depend on age

	Literacy	
T-shirt	Yes	No
Small	0.32	0.68
Large	0.35	0.65

		Literacy	
Age	T-shirt	Yes	No
<5	Small	0.3	0.7
<5	Large	0.3	0.7
>5	Small	0.4	0.6
>5	Large	0.4	0.6

Bayesian Networks

- Bayesian networks use graphs to capture these statement of conditional independence.
- A Bayesian network (BBN) is defined by a graph:
  - Nodes are stochastic variables.
  - Links are dependencies.
    - *Absence of a link denotes independence given a parent.*
- There are two components in a BBN:
  - Qualitative graphical structure.
  - Quantitative assessment of probabilities.



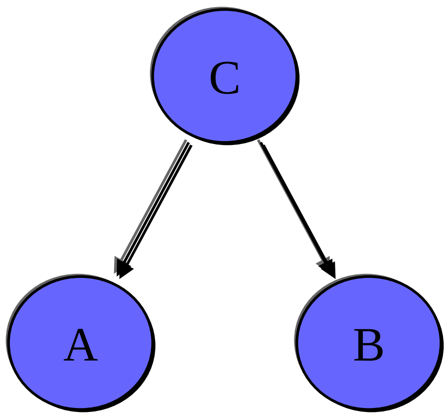
Decomposition

- The power of BBNs is to decompose the joint probability distribution using the graphical structure of conditional independence.

$$p(A|B \wedge C)=p(A|C) \equiv p(A \wedge B\ |C)=p(A|C) \times p(B|C)$$

- Therefore, the graphical structure factorizes the joint probability distribution:

$$p(A \wedge B \wedge C)=p(A|C) \times p(B|C) \times p(C)$$



Example

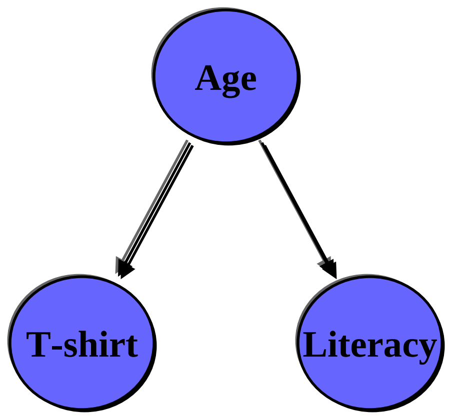
Background knowledge: General rules of behavior.

p(Age=<5)=0.3
p(T-shirt=small| Age=<5)=0.5
p(T-shirt=small|Age=>5)=0.3
p(Literacy=yes|Age=>5)=0.6
p(Literacy=yes|Age=<5)=0.2

Problem: Observation *p(T-shirt=small)*

Solution: The posterior probability distribution of the unobserved nodes given problem: *p(Literacy| T-shirt=small)* and *p(Age| T-shirt=small)*

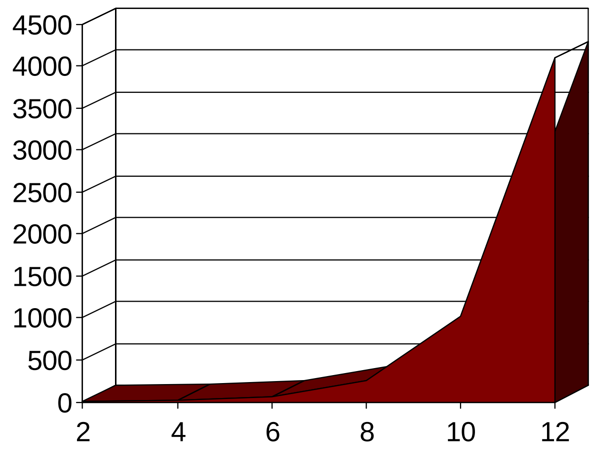
p(Age=<5,T-shirt=small,Literacy=yes)
p(Age=<5,T-shirt=small,Literacy=no)
p(Age=<5,T-shirt=large,Literacy=yes)
p(Age=<5,T-shirt=large,Literacy =no)
p(Age=>5,T-shirt=small,Literacy=yes)
p(Age=>5,T-shirt=small,Literacy=no)
p(Age=>5,T-shirt=large,Literacy=yes)
p(Age=>5,T-shirt=large, Literacy=no)



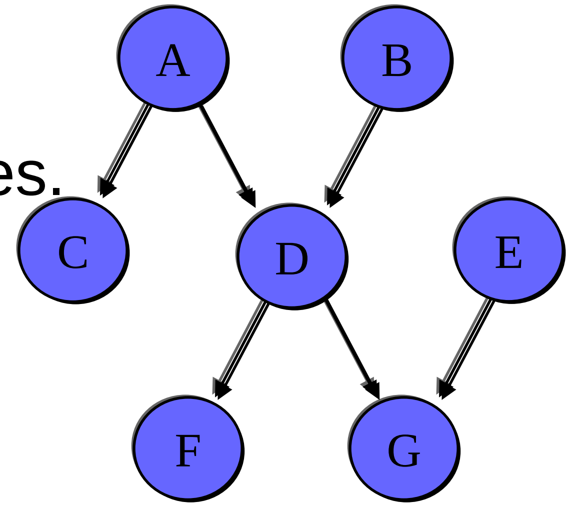
Brute Force

- The brute force solution:
  - Compute the Joint Probability Distribution:
    p(a,b,c,d,e,f,g)= p(a)p(b)p(c|d)p(d|a,b)p(e)p(f|d)p(g|d,e)
  - Marginalize out the variable of interest:

    $$p(d)=\Sigma\ p(a,b,c,e,f,g)$$
  - ✓ Note we have replace $\wedge$ with ,
  - ✓ Cost: we need to sum 2^n probabilities (2^6 = 64).



Knowledge



- Components of a problem:
  Knowledge: graph and probabilities.
  Problem: ε={c and g}.
  Solution: p(d|c,g)=?.

A	p(A)
0	0.3
1	0.7

B	p(B)
0	0.6
1	0.4

E	p(E)
0	0.1
1	0.9

A	B	D	p(D|A,B)
0	0	0	0.40
0	0	1	0.60
0	1	0	0.45
0	1	1	0.55
1	0	0	0.60
1	0	1	0.40
1	1	0	0.30
1	1	1	0.70

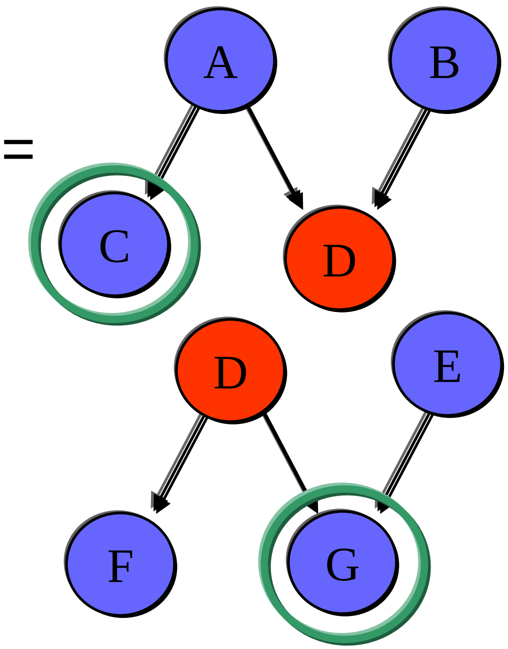
D	E	G	p(G|D,E)
0	0	0	0.90
0	0	1	0.10
0	1	0	0.70
0	1	1	0.30
1	0	0	0.25
1	0	1	0.75
1	1	0	0.15
1	1	1	0.85

A	C	p(C|A)
0	0	0.25
0	1	0.75
1	0	0.50
1	1	0.50

D	F	p(F|D)
0	0	0.80
0	1	0.20
1	0	0.30
1	1	0.70

Method

- In a polytree, each node breaks the graph into two independent graphs and we can deal separatelyε:
  E^+: evidence coming from the parents ($E^+ = \{c\}$).
  E^-: evidence coming from the children ($E^- = \{g\}$).
- Task: p(d|c,g): $p(d| E^+,E^-)$:
  - $p(d| E^+,E^-)= p(E^+,E^-|d)p(d)/p(E^+,E^-)=$
    $=k\ p(E^+|d)p(E^-|d)p(d)= k\ p(E^+,d)p(E^-|d)=$
    $=k\ \lambda_D(d)\pi_D(d)$.
  - $k=1/p(E^+,E^-)$ (Normalizing constant)
    $\pi(d)= p(E^+,d)$ (Parents messages)
    $\lambda(d)= p(E^-|d)$ (Children messages)
  - $p(d| E^+,E^-)=k\ \pi(d)\lambda(d)$



Summary

A				
a	ρ	λ	β	p
0	0.3	1.0	0.3	0.3
1	0.7	1.0	0.7	0.7

B				
b	ρ	λ	β	p
0	0.6	1.0	0.6	0.6
1	0.4	1.0	0.4	0.6

a	ρ	λ
0	0.3	1.0
1	0.7	1.0

a	ρ	λ
0	0.3	1.0
1	0.7	1.0

b	ρ	λ
0	0.6	1.0
1	0.4	1.0

C				
C	ρ	λ	β	p
0	0.425	1.0	0.425	0.425
1	0.575	1.0	0.575	0.575

D				
d	ρ	λ	β	p
0	0.462	1.0	0.462	0.462
1	0.538	1.0	0.538	0.538

E				
b	ρ	λ	β	p
0	0.1	1.0	0.1	0.1
1	0.9	1.0	0.9	0.9

d	ρ	λ
0	0.462	1.0
1	0.538	1.0

d	ρ	λ
0	0.462	1.0
1	0.538	1.0

e	ρ	λ
0	0.1	1.0
1	0.9	1.0

F				
f	ρ	λ	β	p
0	0.531	1.0	0.531	0.531
1	0.469	1.0	0.469	0.469

G				
C	ρ	λ	β	p
0	0.419	1.0	0.419	0.419
1	0.581	1.0	0.581	0.581

Summary

A

a	ρ	λ	β	p
0	0.3	0.42	0.126	0.273
1	0.7	0.48	0.336	0.727

B

b	ρ	λ	β	p
0	0.6	0.540	0.324	0.701
1	0.4	0.345	0.138	0.299

a	ρ	λ
0	0.3	1.0
1	0.7	1.0

a	ρ	λ
0	0.3	0.42
1	0.7	0.48

b	ρ	λ
0	0.6	0.540
1	0.4	0.345

C

C	ρ	λ	β	p
0	0.432	1.0	0.432	0.432
1	0.568	1.0	0.568	0.568

D

d	ρ	λ	β	p
0	1.0	1.0	1.0	1.0
1	0.0	1.0	0.0	0.0

E

b	ρ	λ	β	p
0	0.1	1.0	0.1	0.1
1	0.9	1.0	0.9	0.9

d	ρ	λ
0	1.0	1.0
1	0.0	1.0

d	ρ	λ
0	1.0	1.0
1	0.0	1.0

e	ρ	λ
0	0.1	0.462
1	0.9	0.462

F

f	ρ	λ	β	p
0	0.8	1.0	0.8	0.8
1	0.2	1.0	0.2	0.2

G

C	ρ	λ	β	p
0	0.72	1.0	0.72	0.72
1	0.28	1.0	0.28	0.28

Multiply Connected BBN

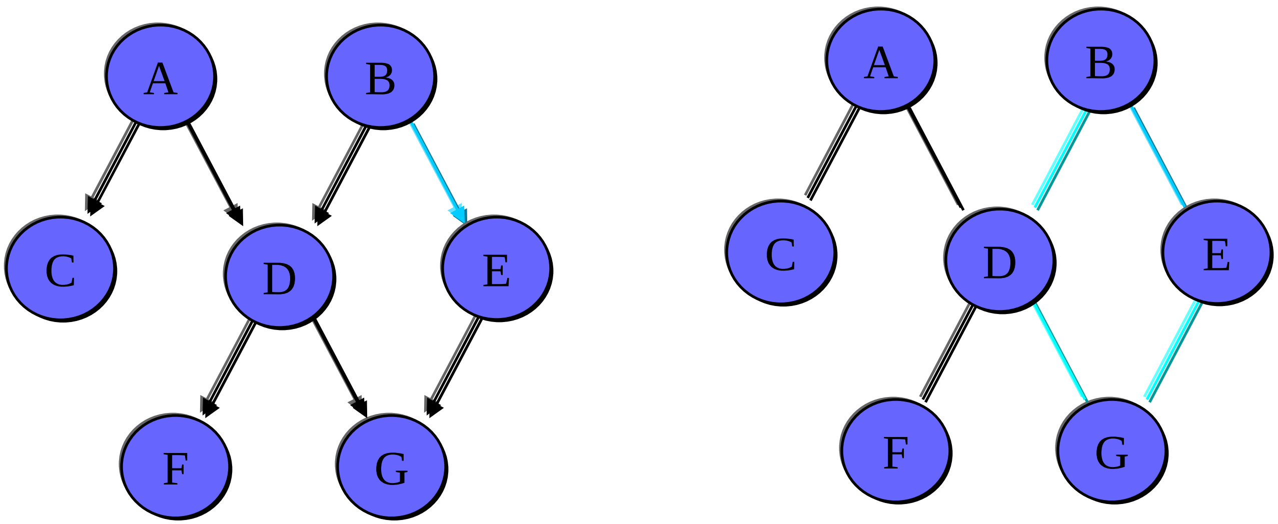
When the BBN is a Multiply connected graph

The associated undirected graph contains a loop.

Each node does not break the network in two parts.

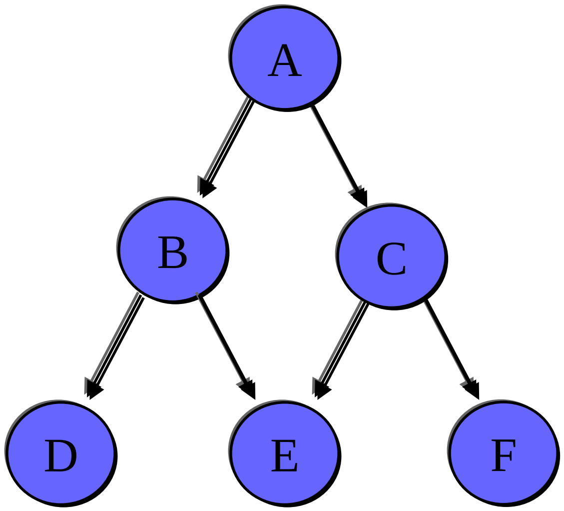
Information may flow through more than one paths.

Pearl’s Algorithm is no longer applicable.



Example

A Multiply connected BBN



A	p(A)
0	0.3
1	0.7

A	B	p(B|A)
0	0	0.4
0	1	0.6
1	0	0.1
1	1	0.9

A	C	p(C|A)
0	0	0.2
0	1	0.8
1	0	0.50
1	1	0.50

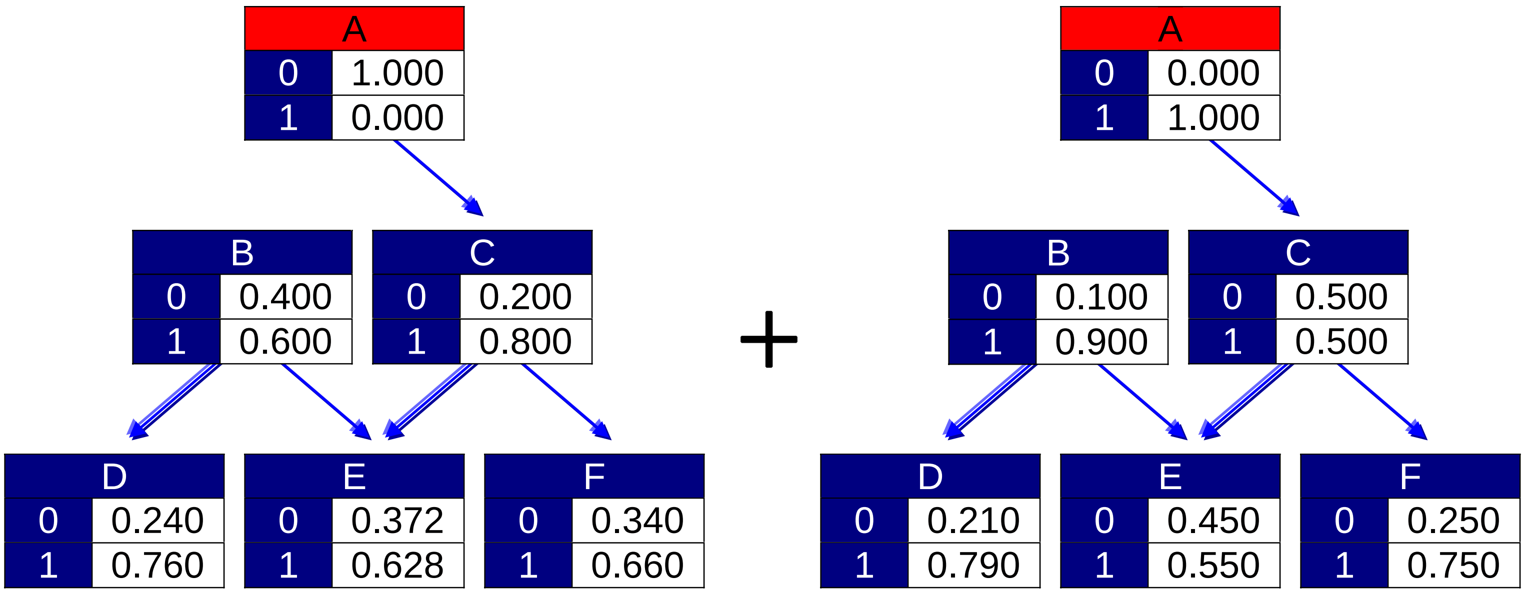
B	D	p(D|B)
0	0	0.3
0	1	0.7
1	0	0.2
1	1	0.8

C	F	p(F|C)
0	0	0.1
0	1	0.9
1	0	0.4
1	1	0.6

B	C	E	p(E|B,C)
0	0	0	0.4
0	0	1	0.6
0	1	0	0.5
0	1	1	0.5
1	0	0	0.7
1	0	1	0.3
1	1	0	0.2
1	1	1	0.8

Conditioning

- Loop cutset: {A}.
- p(B=0)=p(B=0|A=0)p(A=1) +p(B=0|A=1)p(A=1).



A	
0	0.300
1	0.700

B	
0	0.190
1	0.810

C	
0	0.410
1	0.590

D	
0	0.219
1	0.781

E	
0	0.427
1	0.573

F	
0	0.277
1	0.723

Decision Problems

- A Decision Problem has three components:
  - ✓ A set of chance variables.
  - ✓ A set of possible alternative decisions.
  - ✓ A utility function ranking the possible outcomes.
- A set of possible decisions is called a strategy
  - ☞An antibiotic is given together vitamins: dosage of antibiotics is one decision, dosage of vitamin another, the strategy identifies the two dosages together.
- The solution is the strategy that maximizes the expected (value of the) utility. This is called Maximum Expected Utility (MEU) Principle.

Expected Value

- When values are numerical, a BBN can be used to predict the expected value given the evidence:

$$E(A)=\Sigma_i (p(A=a_i) \times a_i)$$

that is, the sum of each possible value a_i of A times its probability $p(A=a_i)$ of being assigned to A:

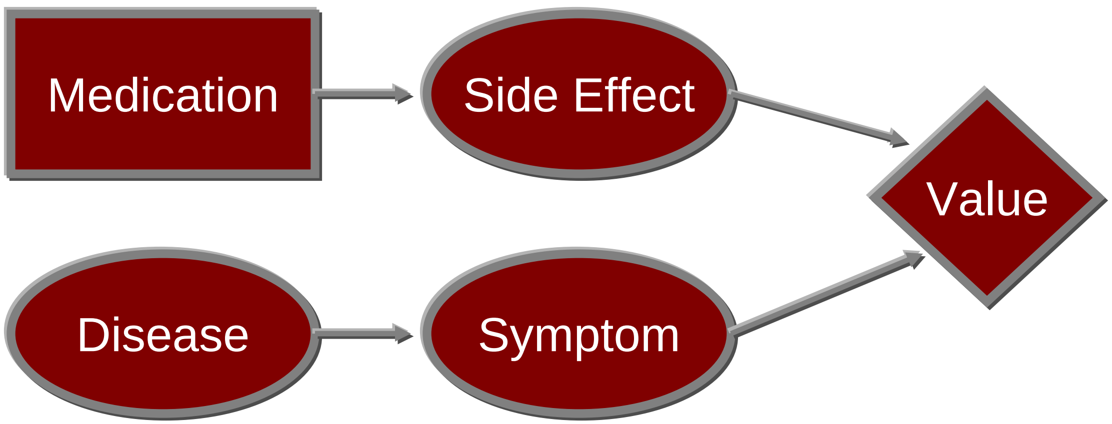
1. Propagate the evidence in the BBN.
2. Apply the formula to calculate expected value.

A	p(A)
0	0.2
1	0.1
2	0.3
3	0.1
4	0.3

→ $E(A)=(0.2\times 0)+(0.1\times 1)+(0.3\times 2)+(0.1\times 3)+(0.3\times 4)=2.2$

Influence Diagrams

- Influence diagram are BBN with 3 kinds of nodes:
  - Chance nodes: stochastic variables (oval)
  - Decision nodes: variables to be set the value (square).
  - Utility nodes: variables ranking the outcomes.

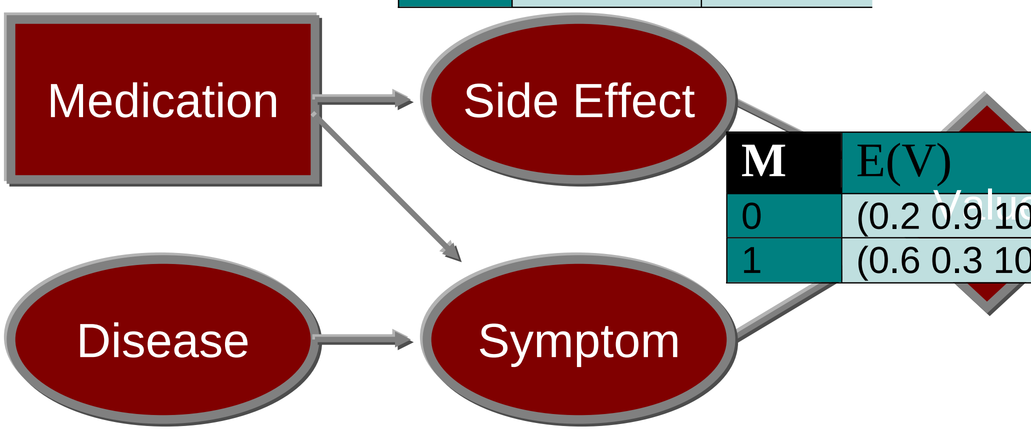


Sy	SE	V
0	0	100
0	1	80
1	0	10
1	1	-100

Making Decisions

- The solution of a decision problem is the decision that maximizes the expected utility, and expected utility is the expected value of a utility node.

M	p(SE=0)	p(SE=1)
0	0.9	0.1
1	0.3	0.7



M	E(V)
0	(0.2 0.9 100)+(0.2 0.1 80)+(0.8 0.9 10)+(0.8 0.1 -100)
1	(0.6 0.3 100)+(0.6 0.7 80)+(0.4 0.3 10)+(0.4 0.7 -100)

M	p(Sy=0)	p(Sy=1)
0	0.2	0.8
1	0.6	0.4

Sy	SE	V
0	0	100
0	1	80
1	0	10
1	1	-100