Artificial & Computational Intelligence

DSE CLZG557

M5 : Probabilistic Representation & Reasoning

Raja vadhana P

Assistant Professor,

BITS - CSIS

BITS Pilani

Pilani Campus

Course Plan

M1 Introduction to AI

M2 Problem Solving Agent using Search

M3 Game Playing, Constraint Satisfaction Problem

M4 Knowledge Representation using Logics

M5 Probabilistic Representation and Reasoning

M6 Reasoning over time, Reinforcement Learning

Module 5:
Probabilistic Representation and Reasoning

innovate achieve lead

A. Inference using full joint distribution

B. Bayesian Networks

I. Knowledge Representation

II. Conditional Independence

III. Exact Inference

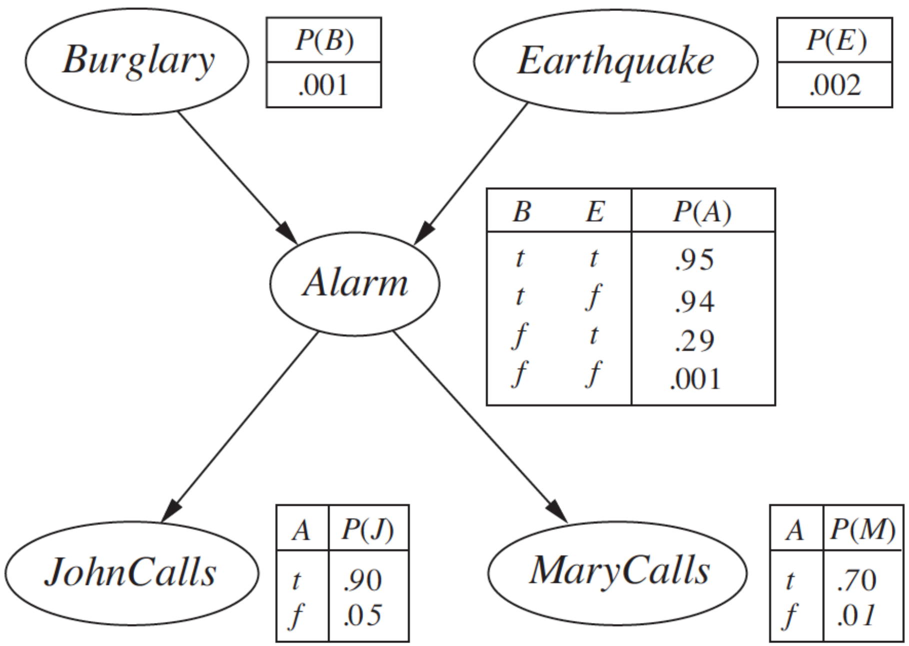
IV. Introduction to Approximate Inference

Reasoning under uncertainty

innovate achieve lead

Building a Bayesian Network

Example Bayesian Net #2



$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

Example Bayesian Net #3

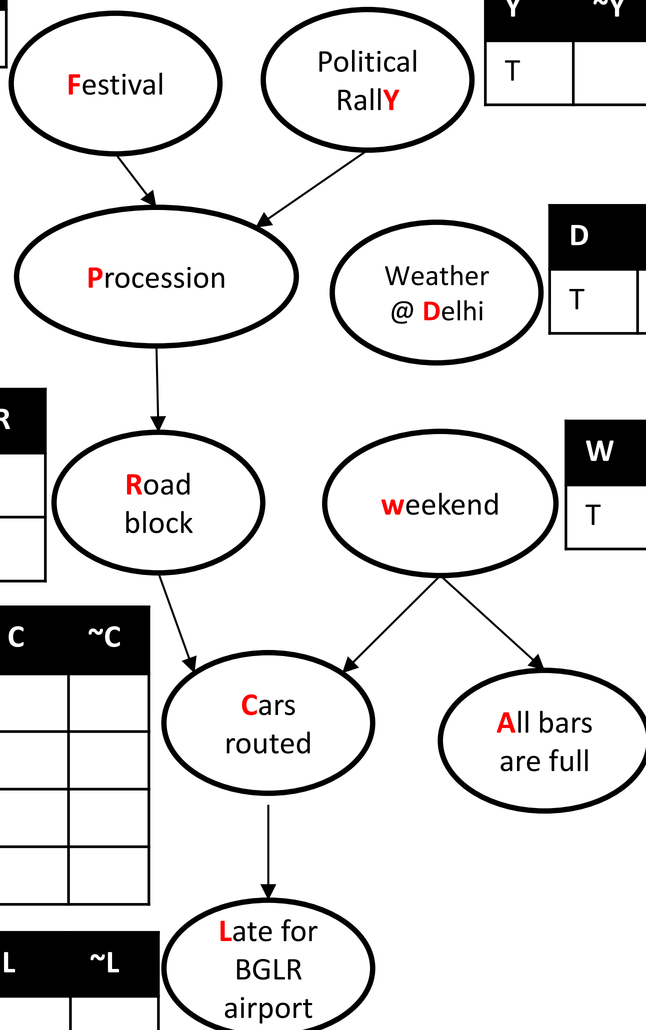
F	~F
T	

F	Y	P	~P
T	T		
T	F		
F	T		
F	F		

P	R	~R
T		
F		

R	W	C	~C
T	T		
T	F		
F	T		
F	F		

C	L	~L
T		
F		



Y	~Y
T	

D	~D
T	

W	~W
T	

W	A	~A
T		
F		

Assumption Bayesian Nets

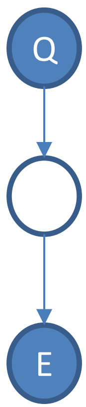
- A node is conditionally independent of its non-descendants given its parents
- A node is conditionally independent of all other nodes in the net , given its parents, children and children’s parents.

Inferences in Bayesian Nets

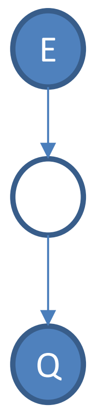
Enumeration

Belief Nets

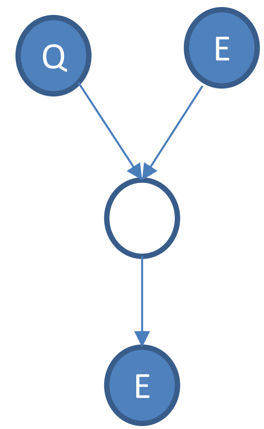
Diagnostic



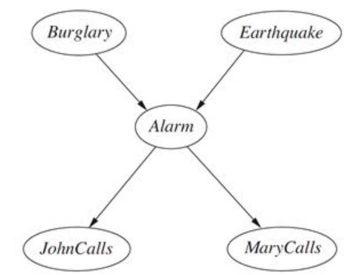
Causal

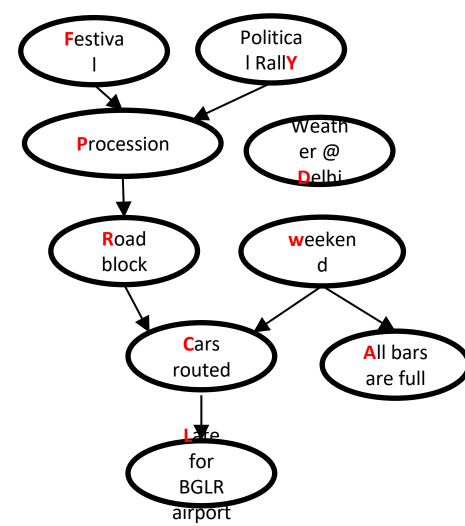


Inter-Casual

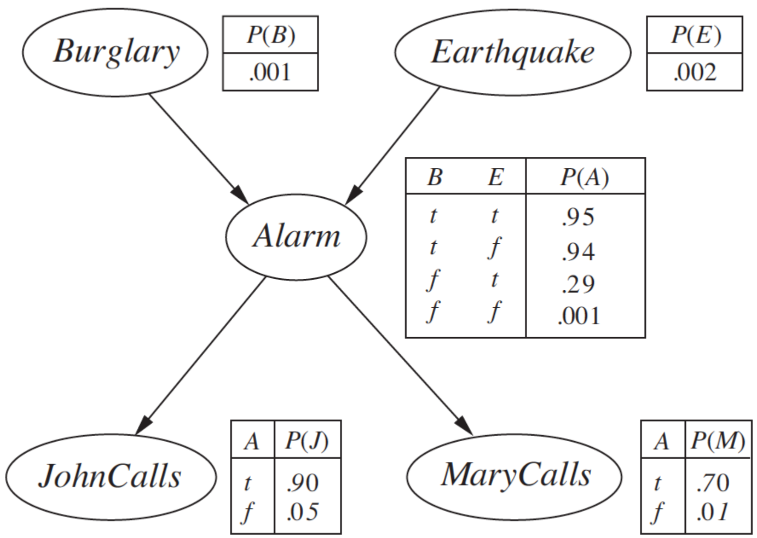


Mixed Inferences





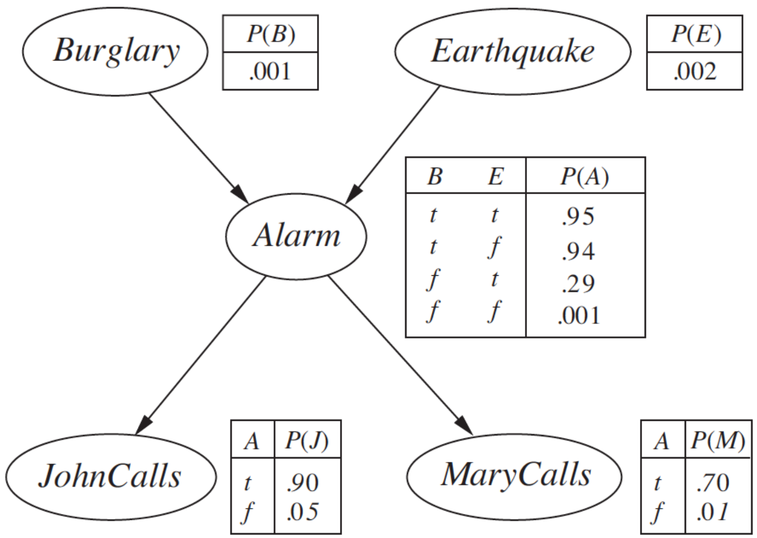
Examples



1. Calculate the probability that alarm has sounded, but neither burglary nor earthquake happened, and both John and Mary called

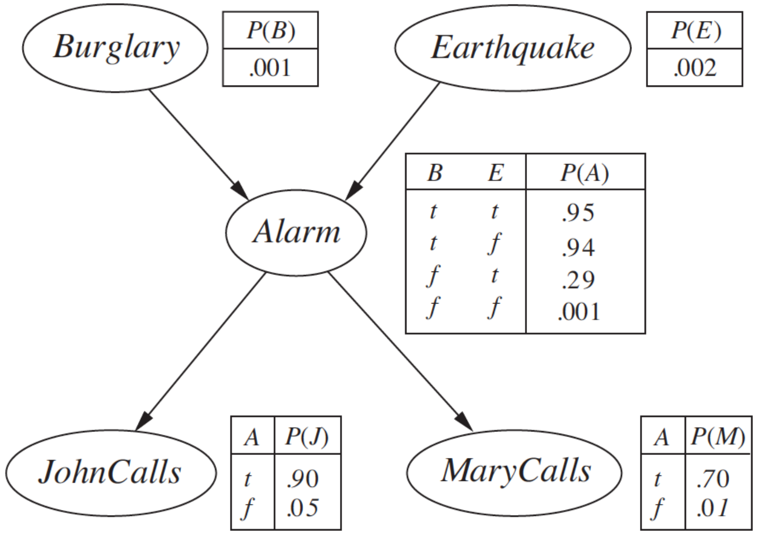
2. What is the probability that Burglary happened given John & Mary called the police

3. What is the probability that John calls given earthquake occurred?



1. Calculate the probability that alarm has sounded, but neither burglary nor earthquake happened, and both John and Mary called

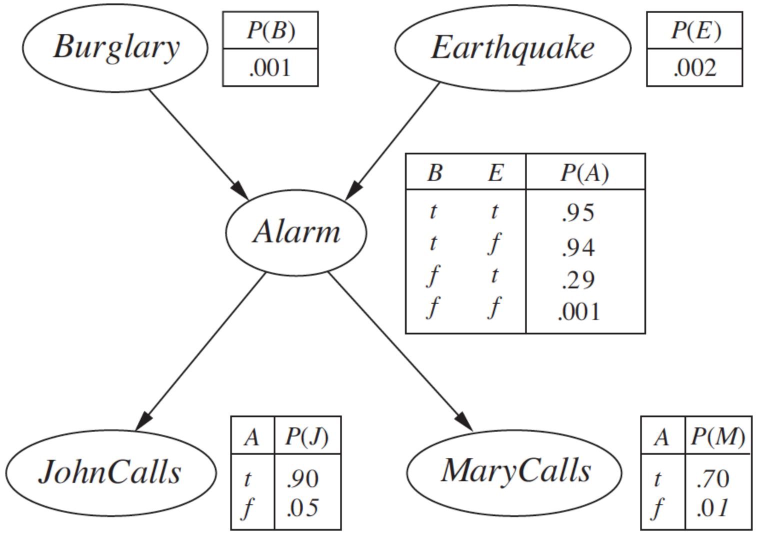
$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j \mid a)P(m \mid a)P(a \mid \neg b \wedge \neg e)P(\neg b)P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$



2. What is the probability that Burglary happened given John & Mary called the police

$$P(B \mid J,M) = \frac{P(B, J, M)}{P(J, M)}$$

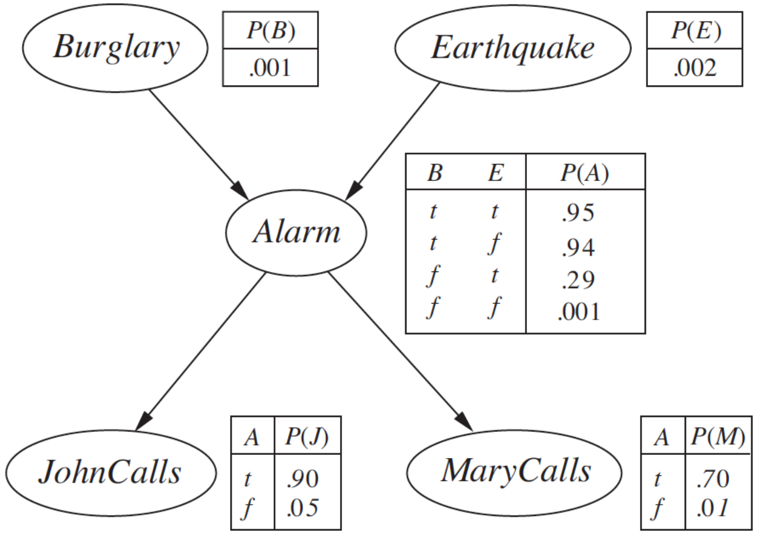
$$P(B \mid J,M) = \frac{\sum_{A, E} P(J, M, A, B, E)}{\sum_{A, B, E} P(J, M, A, B, E)}$$



2. What is the probability that Burglary happened given John & Mary called the police

$$P(B \mid J,M) = \frac{P(B, J, M)}{P(J, M)}$$

$$P(B \mid J,M) = \frac{\sum_{A, E} P(J, M, A, B, E)}{\sum_{A, B, E} P(J, M, A, B, E)}$$



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

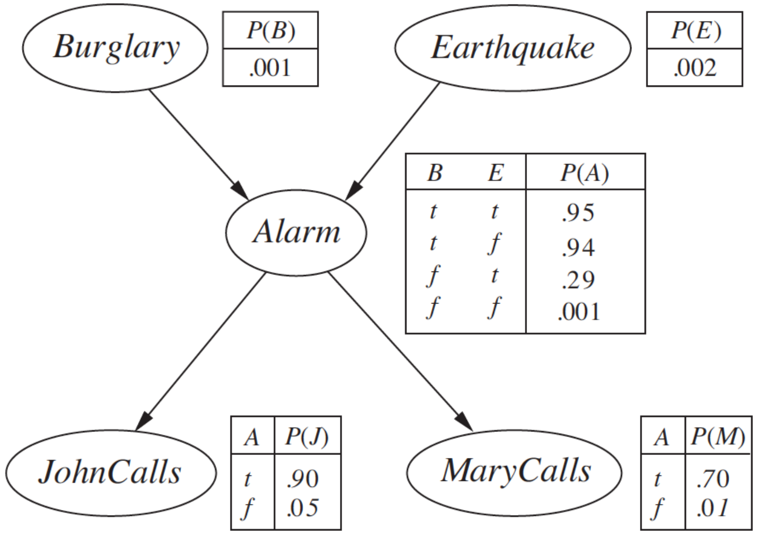
A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

2. What is the probability that Burglary happened given John & Mary called the police

$$P(B \mid J,M) = \frac{P(B, J, M)}{P(J, M)}$$

$$P(B \mid J,M) = \frac{\sum_{A, E} P(J, M, A, B, E)}{\sum_{A, B, E} P(J, M, A, B, E)}$$



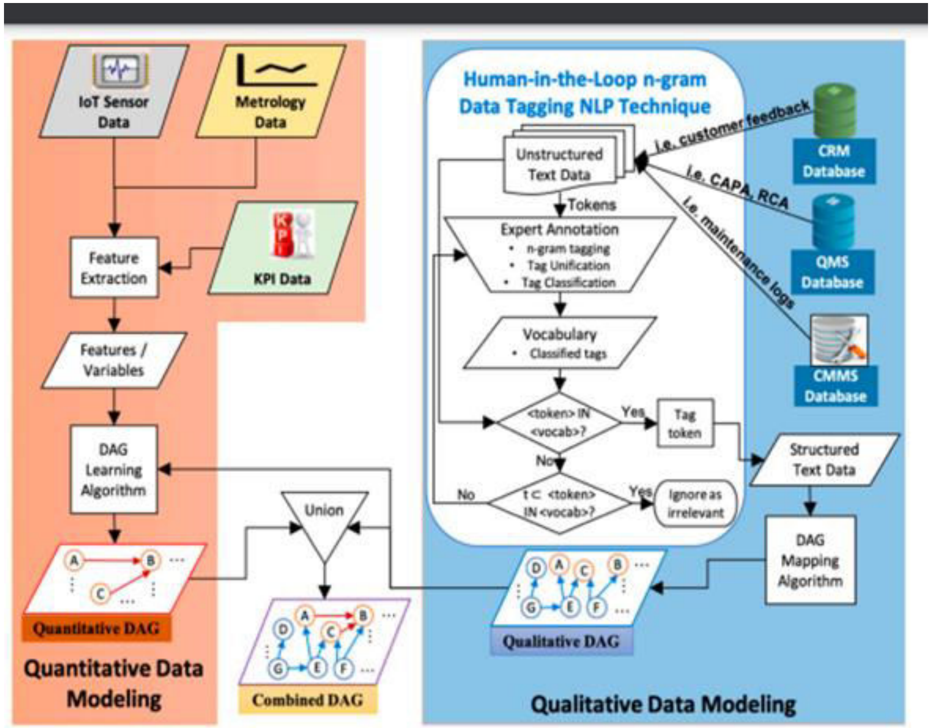
3. What is the probability that John calls given earthquake occurred?

$$P(J \mid E) = \frac{P(J, E)}{P(E)}$$

$$P(J \mid E) = \frac{\sum_{M, A, B} P(J, M, A, B, E)}{\sum_{J, M, A, B} P(J, M, A, B, E)}$$

Bayesian Network

Fault Diagnostic System

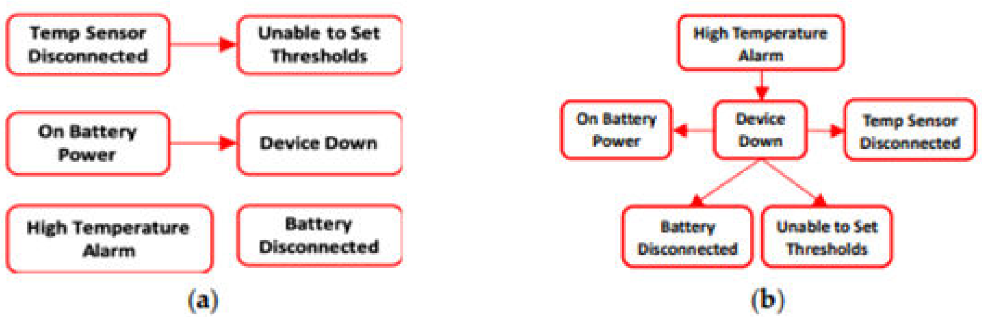


Source Credit : [Sensors 2021 : Fusion-Learning of Bayesian Network Models for Fault Diagnostics]

Bayesian Network

Fault Diagnostic System

Raw Data	Short Description			Resolution Notes		
	On battery power			Power outage due to transformer fire		
Classified Tags	**Symptom**		**Cause(s)**		**Link**	
	on_battery_power		power_outage, transformer_fire		due_to	
BN Mapping	**Child Variable**	**Child State**	**Parent Variable**	**Parent State**	**Ancestor Variable**	**Ancestor State**
	on_battery_power	yes	power_outage	yes	transformer	Fire



Source Credit : [Sensors 2021 : Fusion-Learning of Bayesian Network Models for Fault Diagnostics]

Bayesian Network

Fault Diagnostic System

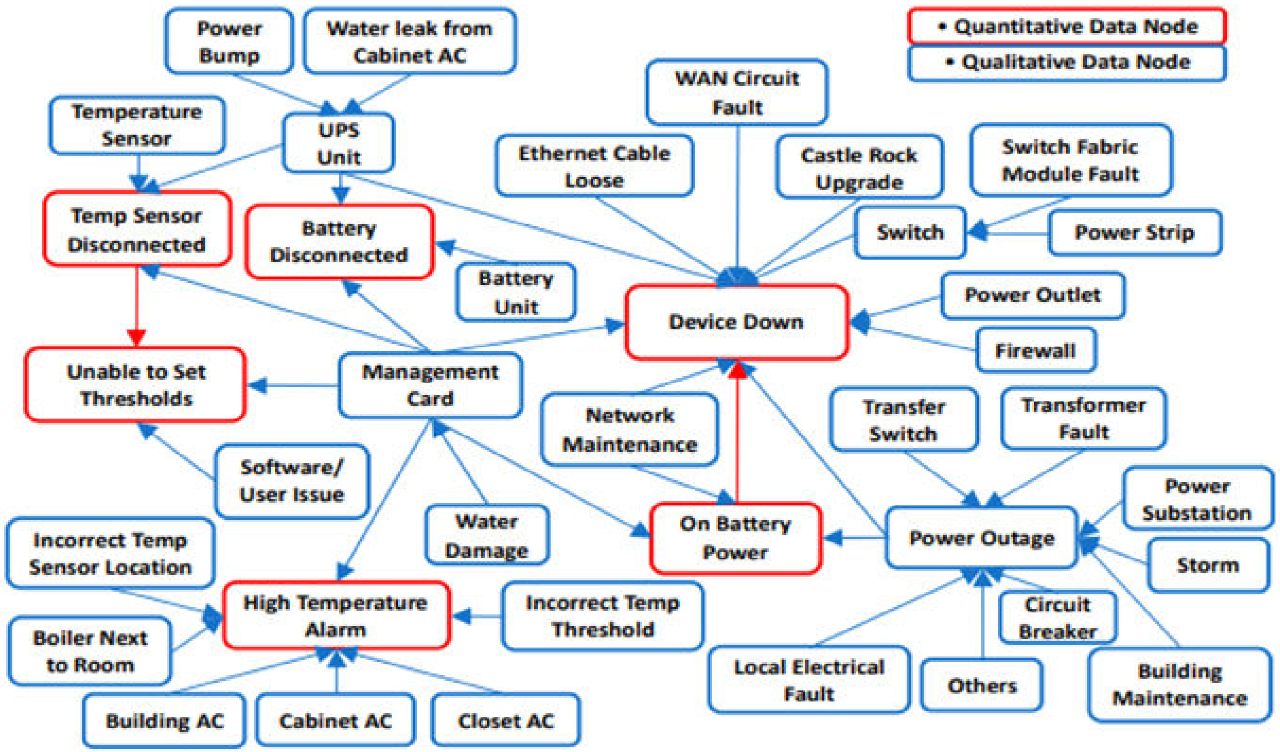
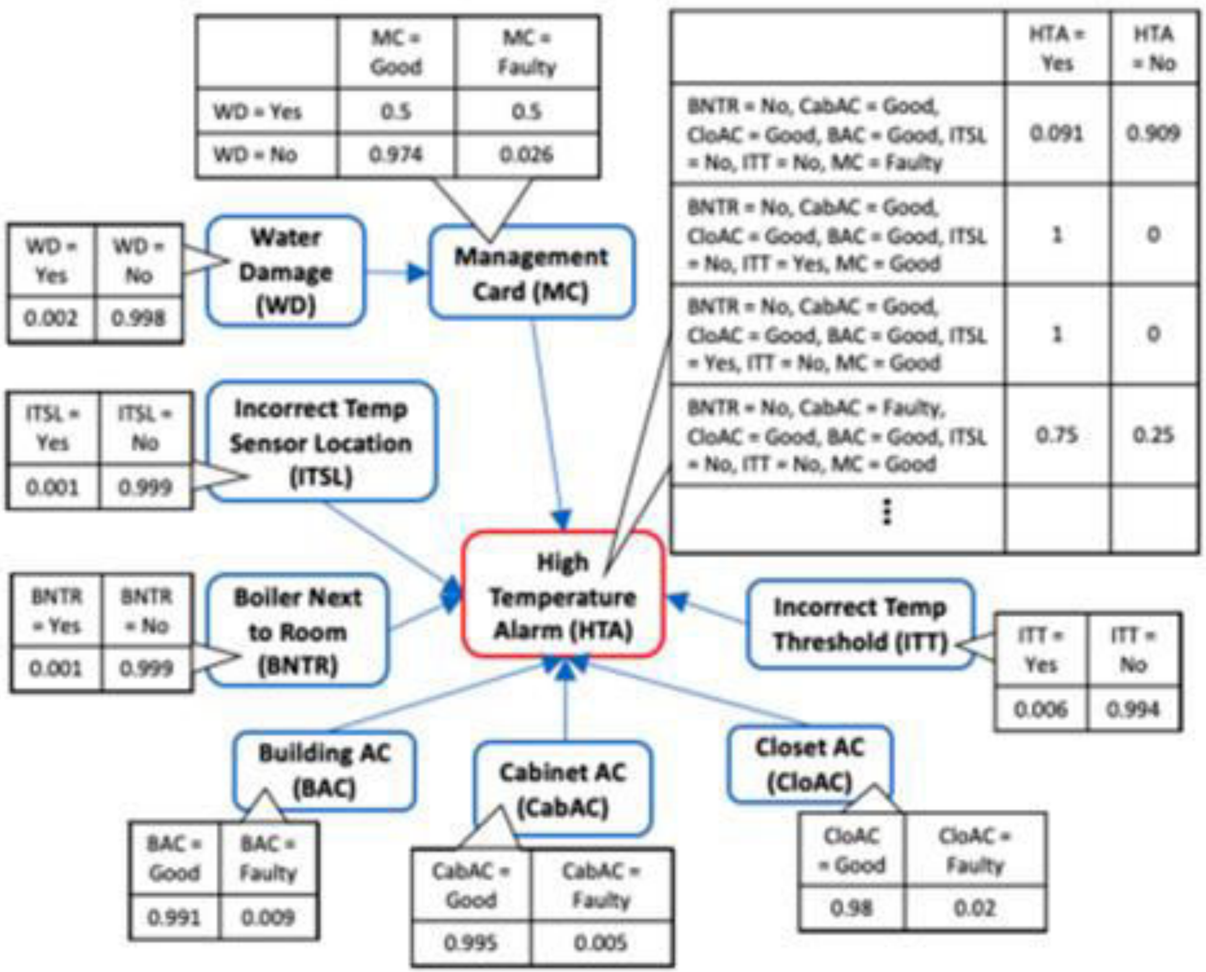


Figure 8. Fused Bayesian Network structure for top six occurring UPS messages.

Source Credit : [Sensors 2021 : Fusion-Learning of Bayesian Network Models for Fault Diagnostics]

Bayesian Network

Fault Diagnostic System

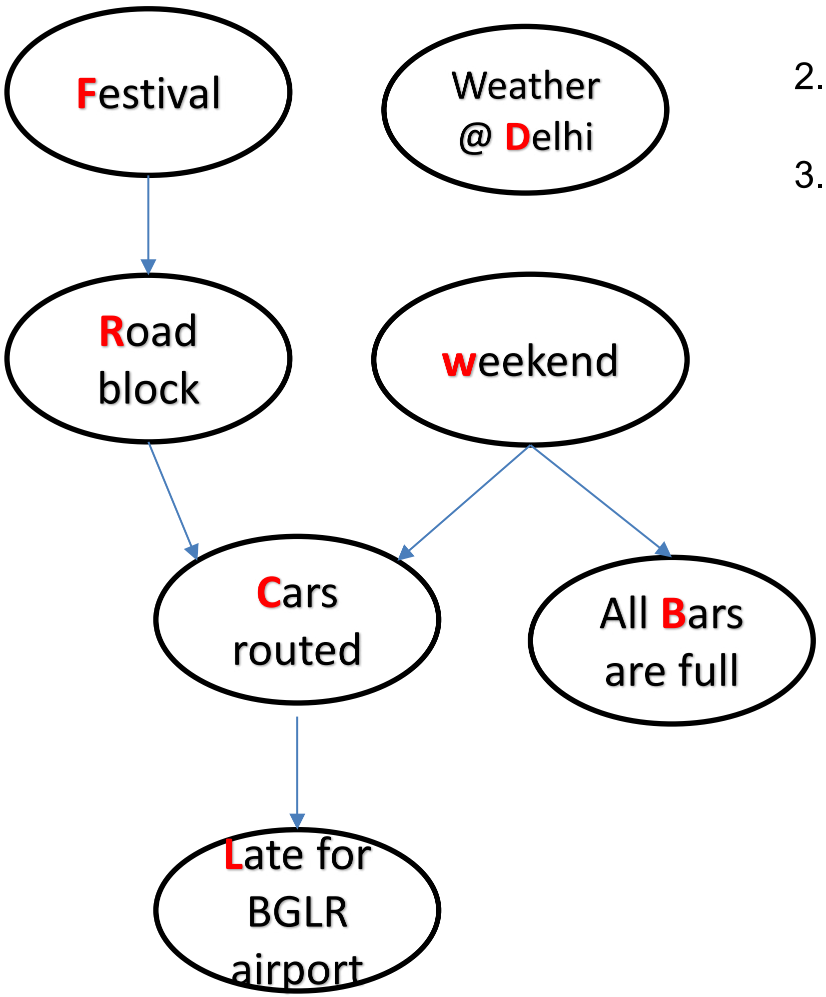


Source Credit : [Sensors 2021 : Fusion-Learning of Bayesian Network Models for Fault Diagnostics]

Inferences in Bayesian Nets

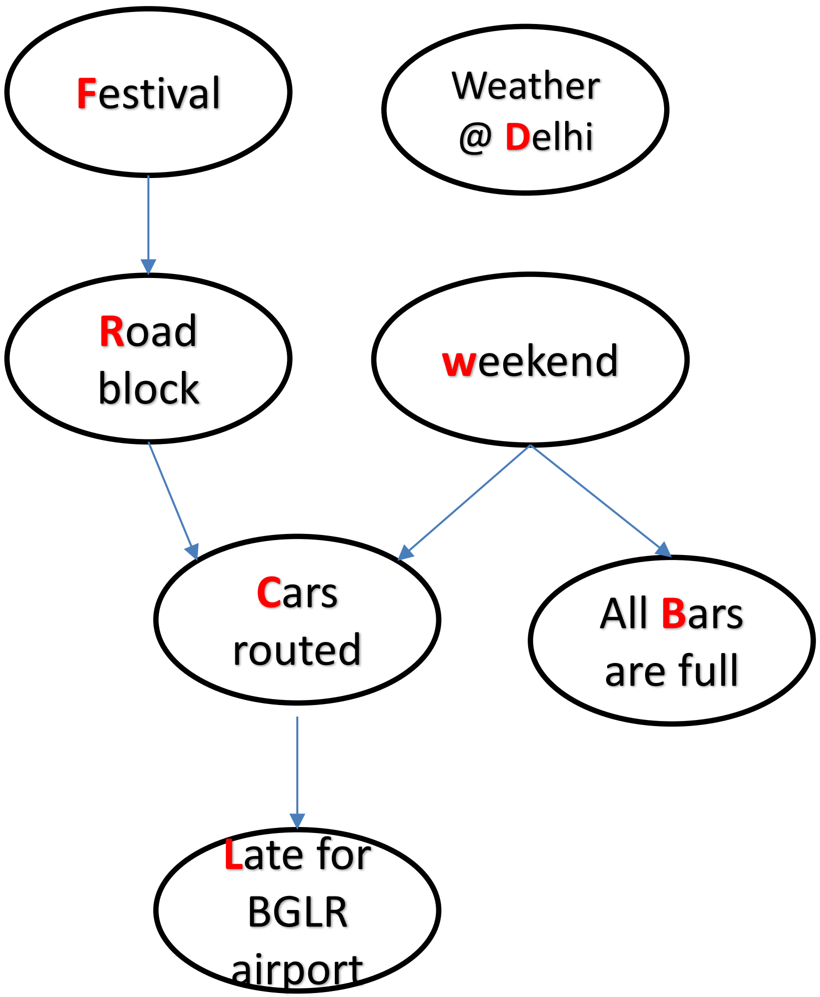
Reduce Guaranteed Independent nodes

D-Connectedness Vs D-Separation



1. Each variable is conditionally independent of its non-descendants, given its parents
2. Eliminate the hidden variables that is neither a query nor an evidence
3. **Two variables are d-separated if they are conditionally independent given evidences**

Try it & Test

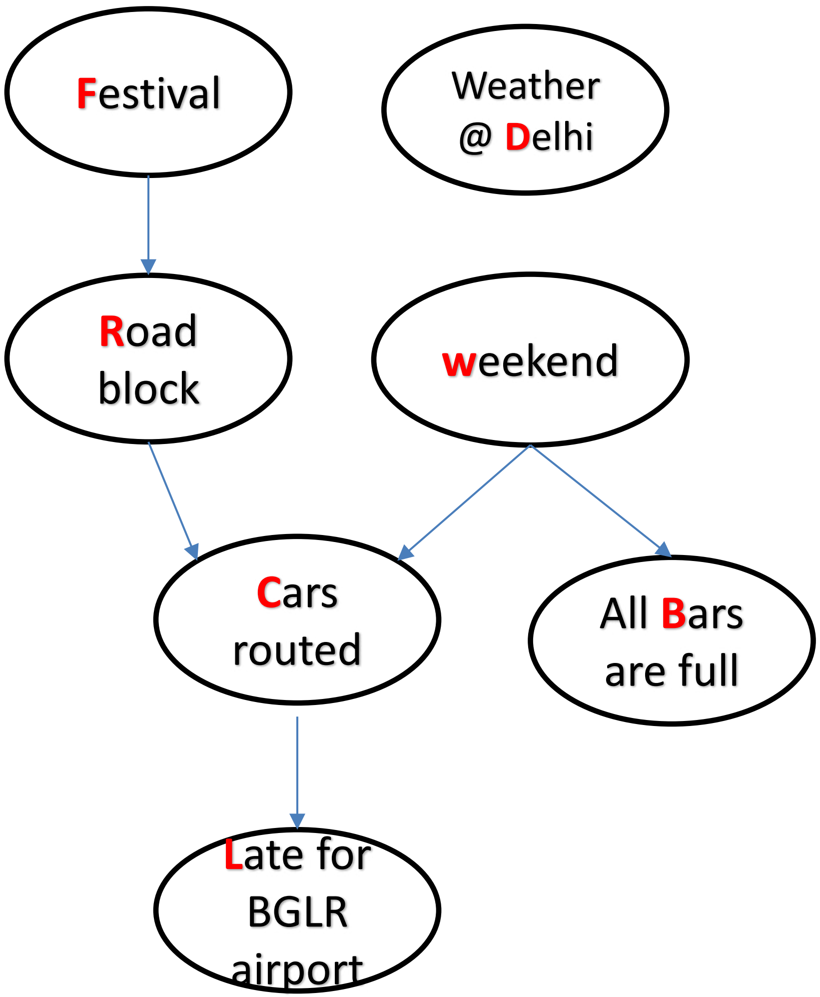


X	Y	Evidence Z	d-sep?
F	W	C	No
L	W	R	No
R	**L**	**C**	**Yes**
B	R	C	No

- $P(R \mid L, C) = P(R \mid L)$

R & L are d-separated ie., conditionally independent given C

D-Separation in Inference



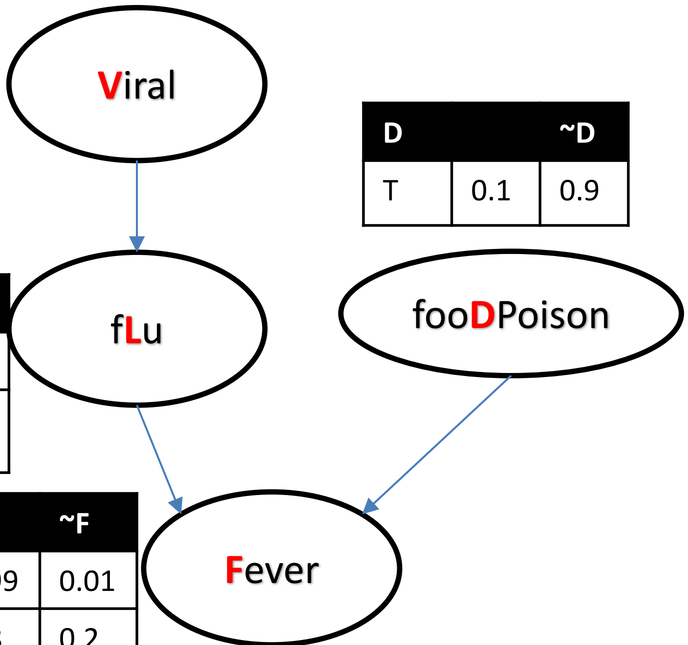
X	Y	Evidence Z	d-sep?
F	W	C	No
L	W	R	No
R	**L**	**C**	**Yes**
B	R	C	No

- P(R | L , C) = P(R| L)

R & L are d-separated ie., conditionally independent given C

D-Separation in Inference

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

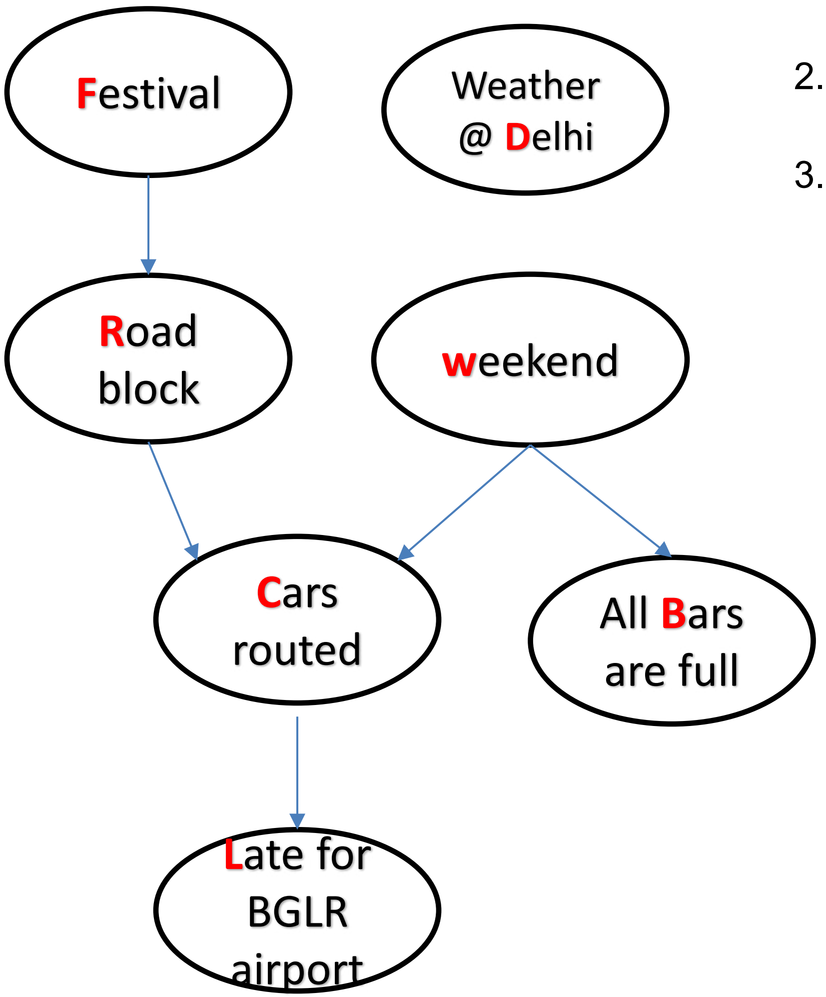
V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

X	Y	Evidence Z	d-sep?
V	F	L	Yes
V	D	L	Yes

- P(V | F , L)
- P(V | D , L)

Variable Elimination



1. Each variable is conditionally independent of its non-descendants, given its parents
2. **Eliminate the hidden variables that is neither a query nor evidence**
3. Two variables are d-separated if they are conditionally independent given evidences

- $P(B) = \sum_{L, B, W, R, F} P(L, C, B, W, R, F)$
  $= \sum_{L} \sum_{B} P(L|C) \, . \, P(B \mid W) \, . \, \sum_{W} P(C \mid W, R) . \sum_{R} P(R \mid F) \, . \, \sum_{F} P(F)$
  $= \mathbf{P(B \mid W)}$

All other variables are hidden w.r.t to B as (L, C, R, F) are neither evidence nor query nor (L, C, R, F) ∊ Ancestors(W, B)

This is variable elimination example targeting irrelevant nodes

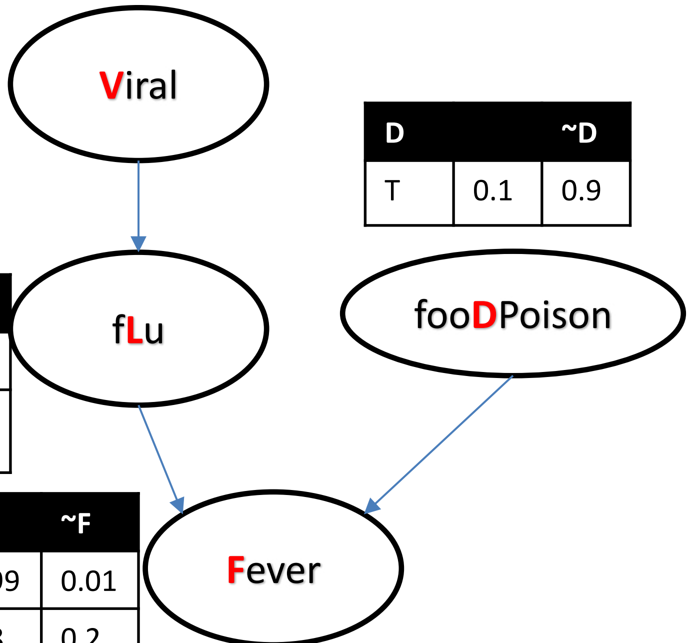
Approximate Inferences in Bayesian Nets

Introduction

Prior Sampling

Sample Generation by Randomization

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

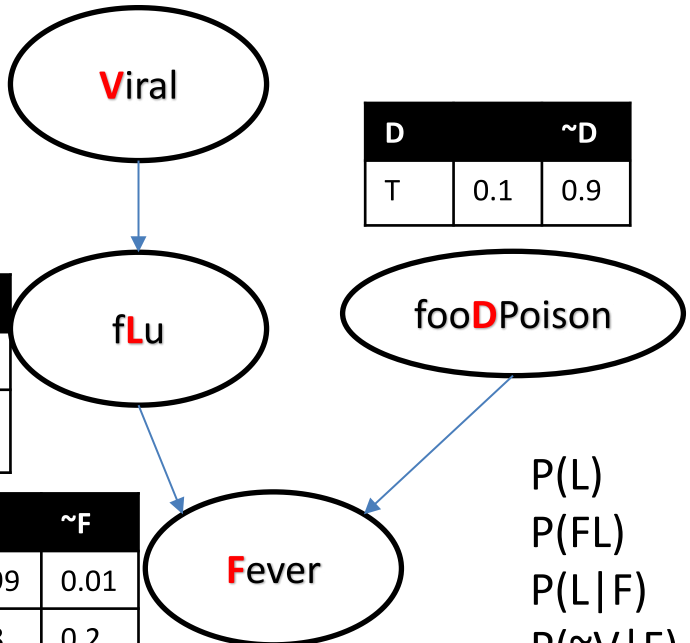
L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

V	L	D	F
T	T	F	T
F	F	F	F
T	F	F	T
F	T	F	T
..			
.....			

0.3, 0.2, 0.6, 0.58, 0.73, 0.87, 0.15, 0.6, 0.57, 0.85, 0.12, 0.004, 0.93 , 0.0002, 0.9,0.55…………..

Prior Sampling

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

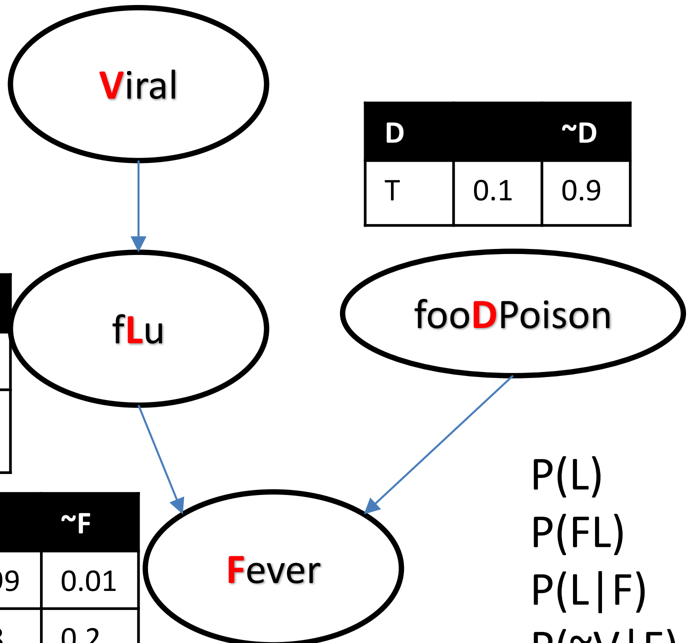
V	L	D	F
T	T	F	T
F	F	F	F
T	F	F	T
F	T	F	T
T	T	F	T
T	F	F	F
F	F	F	T
T	F	F	F

P(L) = 3/8
P(FL) = 3/8
P(L|F) = 3/5
P(~V|F) = 2/5
P(L|V~F) = 0
P(F|D) = ?????

Rejection Sampling

Sample Generation by Randomization

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

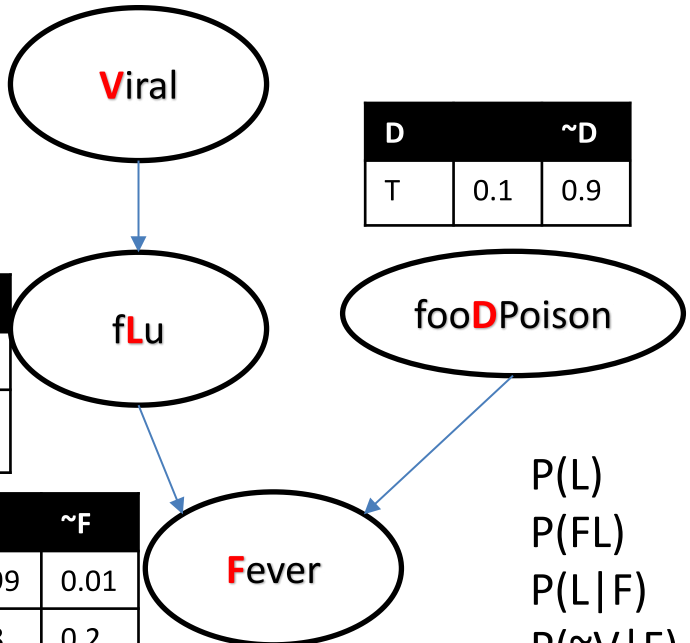
V	L	D	F
T	T	F	
T	T	F	
T	T	F	
F	F	T	F
T	F	T	T
…..			

P(L) = 3/8
P(FL) = 3/8
P(L|F) = 3/5
P(~V|F) = 2/5
P(L|V~F) = 0
P(F|D) = ?????

0.3, 0.2, 0.6, 0.58, 0.73, 0.87, 0.15, 0.6, 0.57, 0.85, 0.12, 0.004, 0.93 , 0.0002, 0.9,0.555,0.38…………..

Rejection Sampling

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

V	L	D	F
T	T	T	T
F	F	T	F
T	F	T	T
F	T	T	T
T	T	T	T
T	F	T	F
F	F	T	T
T	F	T	F

$P(L) = 3/8$

$P(FL) = 3/8$

$P(L|F) = 3/5$

$P(\sim V|F) = 2/5$

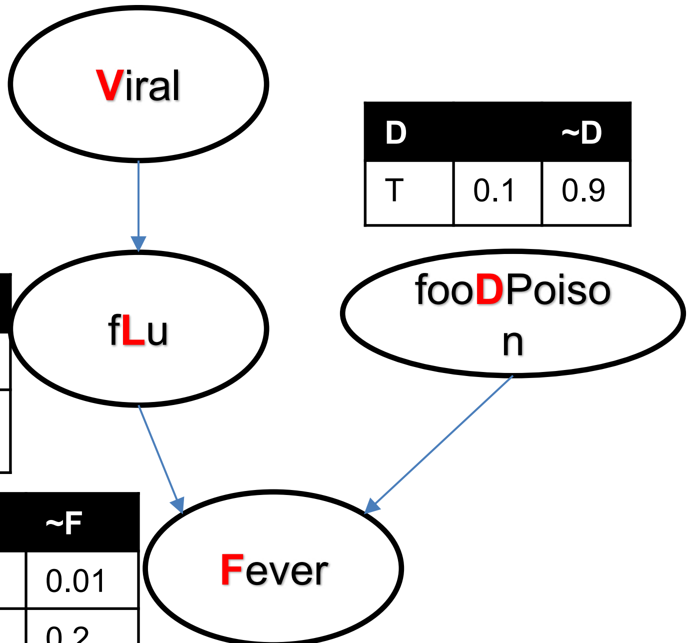
$P(L|V\sim F) = 0$

$P(F|D) = 5/8$

Likelihood Weighing

Sample Generation by Randomization

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

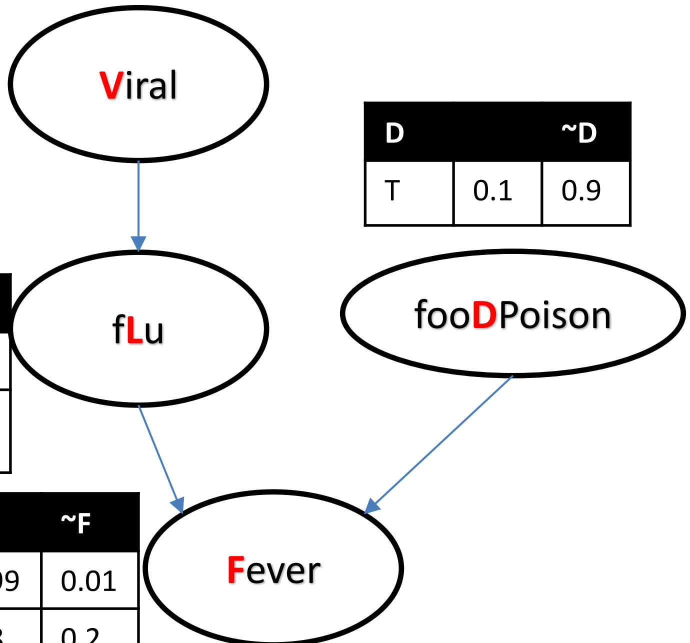
V	L	D	F	wgt
F	F	T	T	0.4*1* 0.1 *1=
F	F	T	T	
F	F	T	T	
F	F	T	T	
F	F	T	T	
F	T	T	T	
F	T	T	F	

P(~F|D,~V) = 0.04 / 7*0.04

0.3, 0.2, 0.58, 0.73, 0.87, 0.15, 0.6, 0.57, 0.85, 0.12, 0.004, 0.93 , 0.0002, 0.99,,………..

Likelihood Weighing

V		~V
T	0.6	0.4



D		~D
T	0.1	0.9

V	L	~L
T	0.8	0.2
F	0.01	0.99

L	D	F	~F
T	T	0.99	0.01
T	F	0.8	0.2
F	T	0.8	0.2
F	F	0.05	0.95

V	L	D	F	wgt
F	F	T	F	1*0.99* 0.1 *1=
F	F	T	T	1*0.99* 0.1 *1=
F	F	T	T	1*0.99* 0.1 *1=
T	F	T	F	1*0.2* 0.1 *1=

P(F|D,~L) = 0.099+0.099 / (3*0.099 + 0.02)

Bayesian Network

Privacy Preserving Data Augmentation / Generation

[0,4], [5,9], …..,[95,99] → (1,6,….,1) n=100 → (1,3,…..,0) n=93
Marginal Distribution(Sensitive Attribute) =(1/93, 3/93……0/93)

Table 2. Example noisy marginal distributions for network $\mathcal{N}_1$.

(a) Marginal distribution for 'age'

'age'	probability
[0, 4]	0.04
[5, 9]	-0.1
…	…
[95, 99]	0.08

(b) Marginal distribution for 'age' and 'higrade'

'age' \ 'higrade'	01	02
[0, 4]	0.02	0.02
[5, 9]	-0.1	0
…	…	…
[95, 99]	0	0

Table 3. Processed noisy marginal distributions for network $\mathcal{N}_1$.

(a) Marginal distribution for 'age'

'age'	probability
[0, 4]	0.05
[5, 9]	0
…	…
[95, 99]	0.1

(b) Marginal distribution for 'age' and 'higrade'

'age' \ 'higrade'	01	02
[0, 4]	0.025	0.025
[5, 9]	0	0
…	…	…
[95, 99]	0	0

Table 4. Noisy marginal distributions for network $\mathcal{N}_1$ that are consistent on attribute 'age'.

(a) Marginal distribution for 'age'

'age'	probability
[0, 4]	0.05
[5, 9]	0
…	…
[95, 99]	0.067

(b) Marginal distribution for 'age' and 'higrade'

'age' \ 'higrade'	01	02
[0, 4]	0.025	0.025
[5, 9]	0	0
…	…	…
[95, 99]	0.0335	0.0335

Source Credit : [TPDP 2020 : Synthetic Data Generation with Differential Privacy via Bayesian Networks]

Required Reading: AIMA - Chapter #13, #14.1, #14.2

Thank You for all your Attention

Note : Some of the slides are adopted from AIMA TB materials

BITS Pilani
Pilani Campus

Artificial & Computational Intelligence

DSE CLZG557

M6 : Reasoning over time & Reinforcement Learning

Raja vadhana P

Assistant Professor,

BITS - CSIS

Course Plan

M1 Introduction to AI

M2 Problem Solving Agent using Search

M3 Game Playing, Constraint Satisfaction Problem

M4 Knowledge Representation using Logics

M5 Probabilistic Representation and Reasoning

M6 Reasoning over time, Reinforcement Learning

Reasoning Over Time

Time & Uncertainty

$$\begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.4 & +ve \\ 0.6 & -ve \end{array}$$

$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.52 & +ve \\ 0.48 & -ve \end{array}$$

$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.466 & +ve \\ 0.534 & -ve \end{array}$$

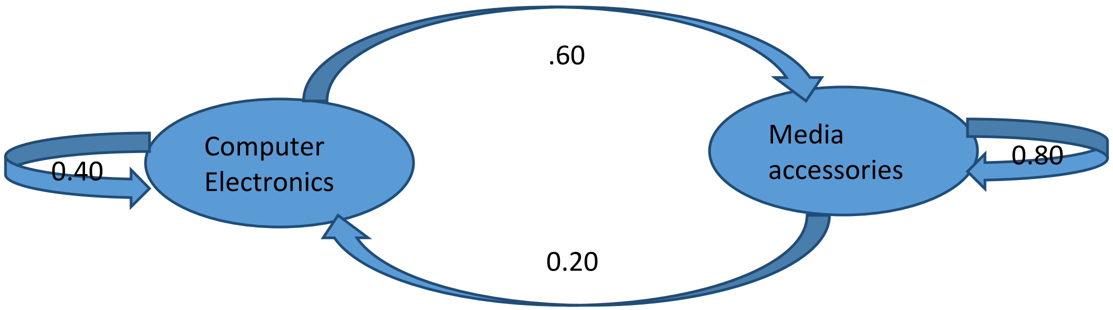
$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.4903 & +ve \\ 0.5097 & -ve \end{array}$$

$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.4903 & +ve \\ 0.5097 & -ve \end{array}$$

$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.48428575 & +ve \\ 0.51571425 & -ve \end{array}$$

$$= \begin{array}{ccc} +ve & -ve & \square \\ 0.25 & 0.7 & +ve \\ 0.75 & 0.3 & -ve \end{array} * \begin{array}{cc} 0.4820714125 & +ve \\ 0.5179285875 & -ve \end{array}$$

Morkov Model



Transition Model

$$\begin{array}{cc} C & M \\ \hline \end{array}$$

$$\begin{array}{c c} \begin{vmatrix} 0.40 & 0.20 \\ 0.60 & 0.80 \end{vmatrix} & \begin{array}{c} C \\ M \end{array} \end{array}$$

innovate achieve lead

Next Session Plan :

- Hidden Morkov Models
- Inferences in Temporal Models