Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 11

Feb, 03, 2021

422 big picture

StarAI (statistical relational AI)

Hybrid: Det +Sto

Prob CFG Parsing

Prob Relational Models

Markov Logics

	Deterministic	Stochastic
Query	*Logics* *First Order Logics* *Ontologies* • Full Resolution • SAT	***Belief Nets*** Approx. : Gibbs ***Markov Chains and HMMs*** Forward, Viterbi…. Approx. : Particle Filtering ***Undirected Graphical Models*** ***Markov Networks*** ***Conditional Random Fields***
Planning		***Markov Decision Processes and Partially Observable MDP*** • Value Iteration • Approx. Inference ***Reinforcement Learning***

Applications of AI

Representation

Reasoning Technique

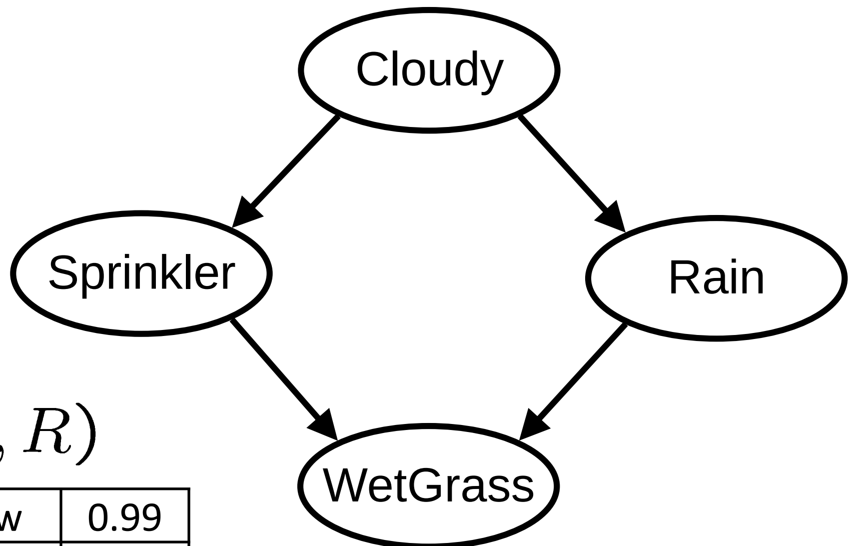
Lecture Overview

- **Recap of BNs** Representation and Exact Inference
- <u>Start Belief Networks **Approx. Reasoning**</u>
    - Intro to **Sampling**
    - First Naïve Approx. Method: **Forward Sampling**
    - Second Method: **Rejection Sampling** *(probably on Fri)*

Bnet basic example: structure + cond. Prob.

$P(C)$

+c	0.5
-c	0.5



$P(S|C)$

+c	+s	0.1
	-s	0.9
-c	+s	0.5
	-s	0.5

$P(R|C)$

+c	+r	0.8
	-r	0.2
-c	+r	0.2
	-r	0.8

$P(W|S,R)$

+s	+r	+w	0.99
		-w	0.01
	-r	+w	0.90
		-w	0.10
-s	+r	+w	0.90
		-w	0.10
	-r	+w	0.01
		-w	0.99

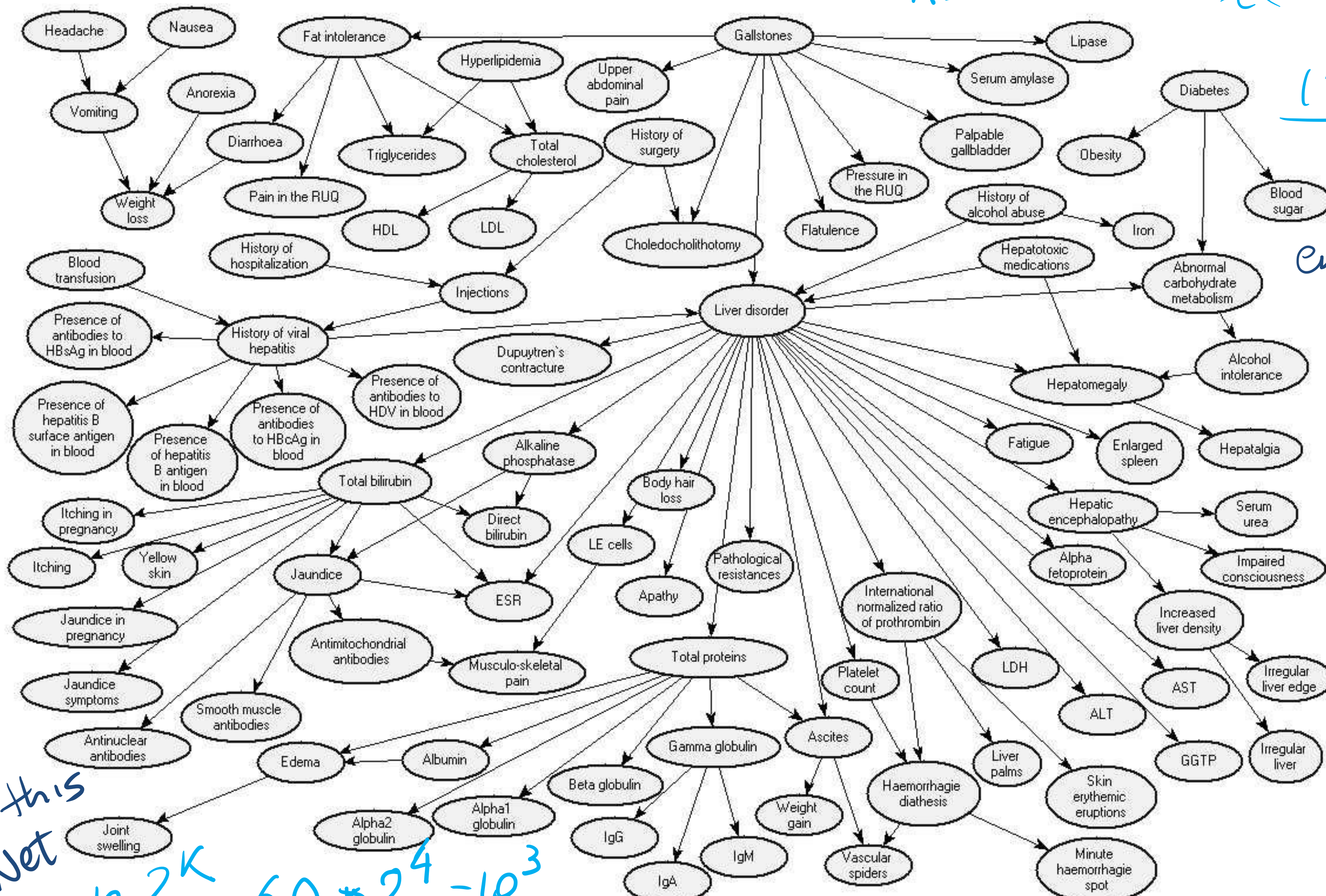
Realistic BNet: Liver Diagnosis

Source: Onisko et al., 1999

~60 nodes

JPD $n = \approx 60$ $\sim 2^{60} = (2^{10})^6$

10^{18} entries

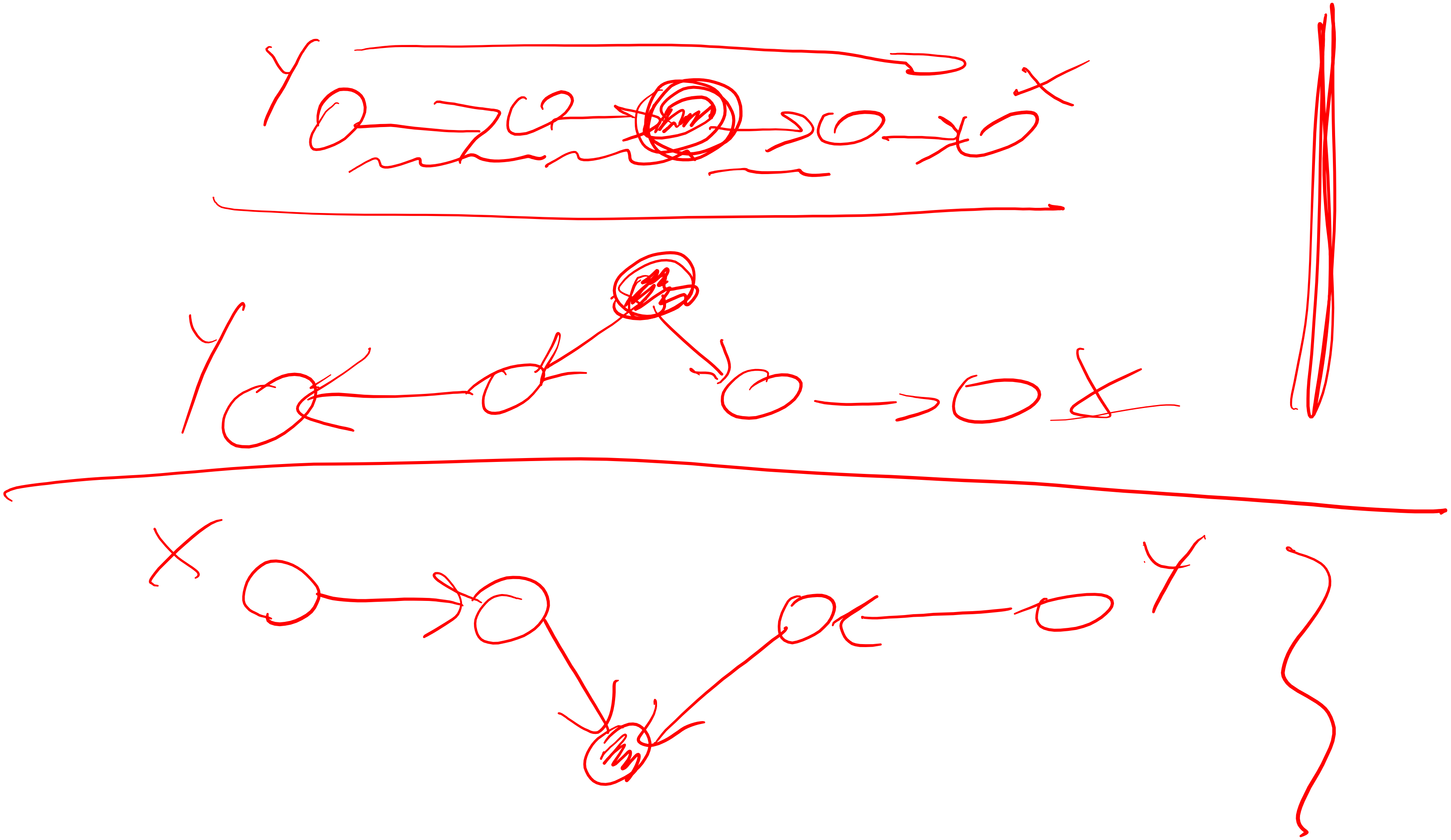


for this BNet

~ 2K

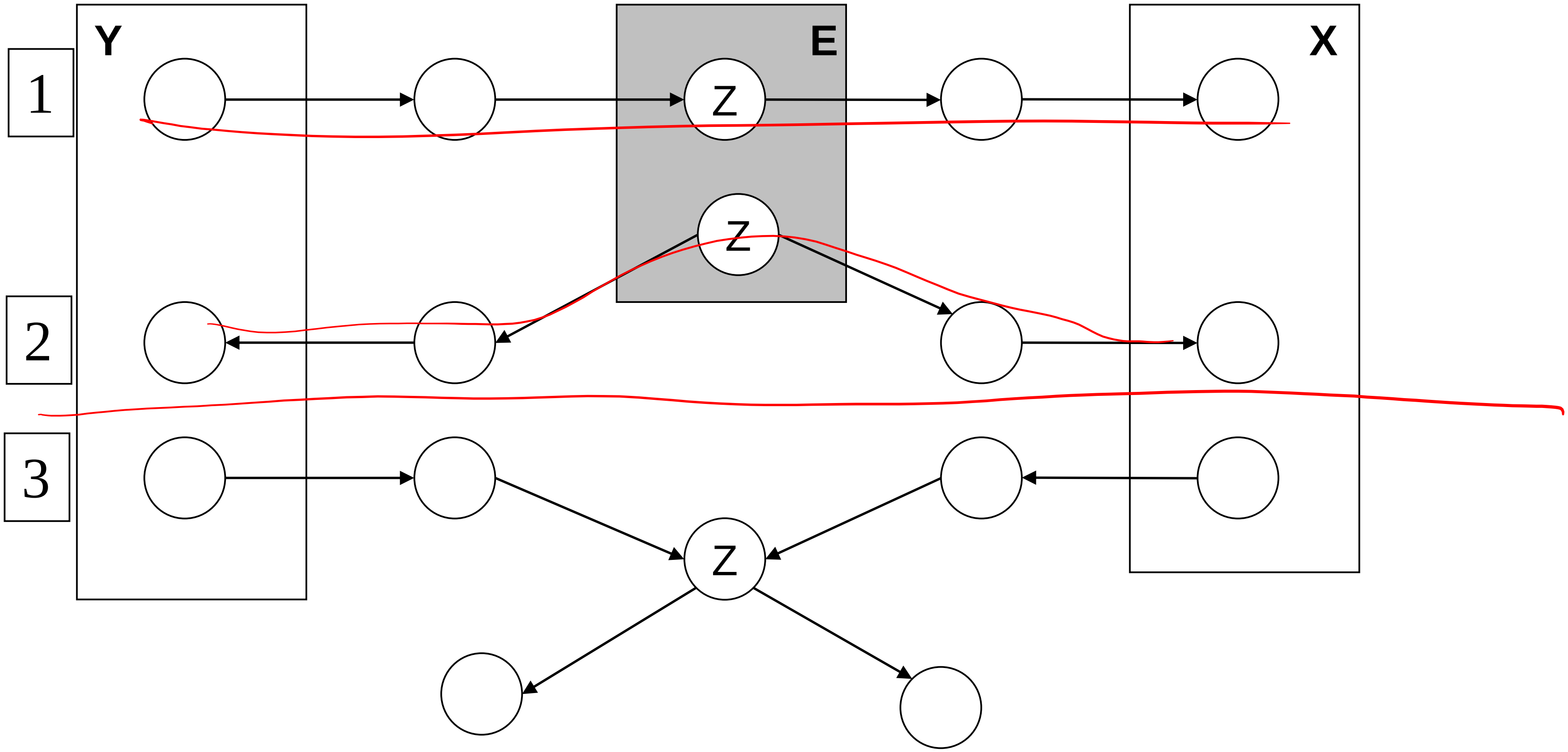
$60 * 2^4 = 10^3$

Revise (in)dependencies……



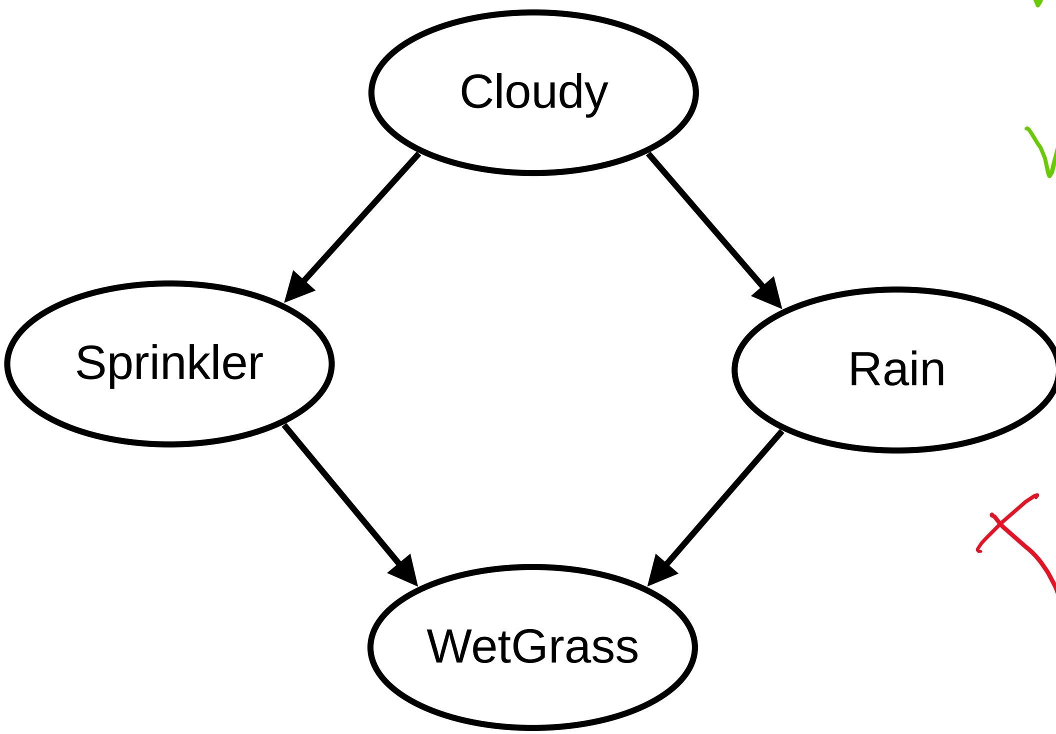
Conditional Independencies

Or, blocking paths for probability propagation. Three ways in which a path between X to Y can be blocked, (1 and 2 given evidence E)



Note that, in 3, X and Y become dependent as soon as I get evidence on Z or on *any of its descendants*

Bnet basic example: independence



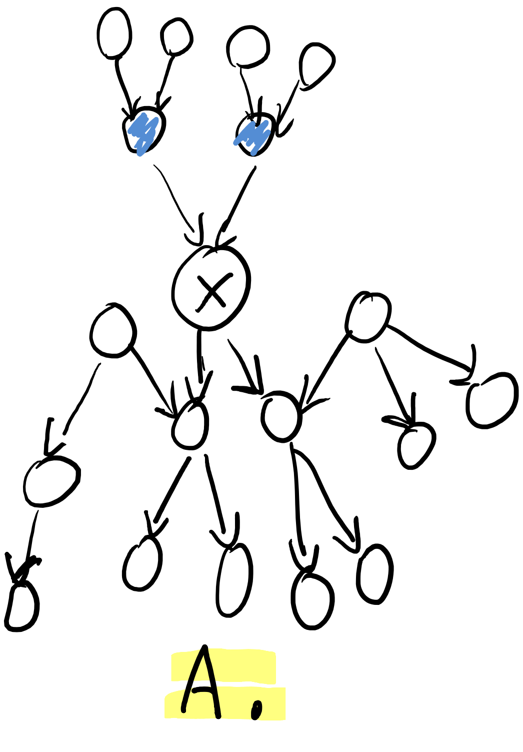
✓ Indep. R, S given C

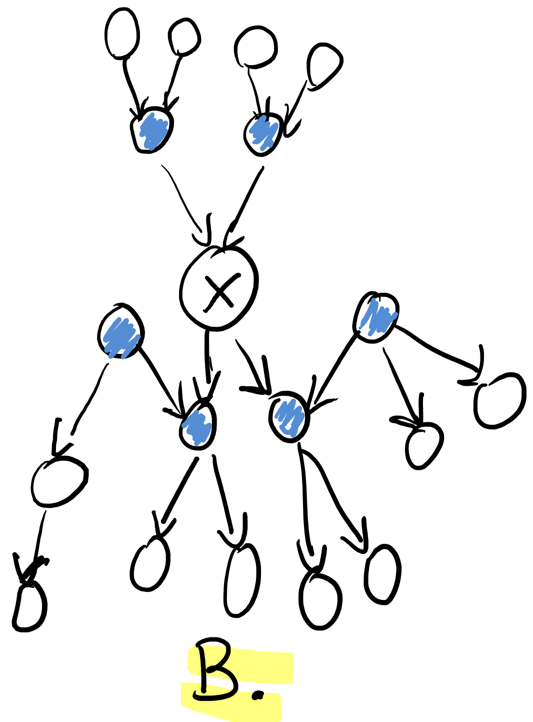
✓ Indep C, W given S, R

✗ Indep. C, W given R

✗ Indep R, S given C, W

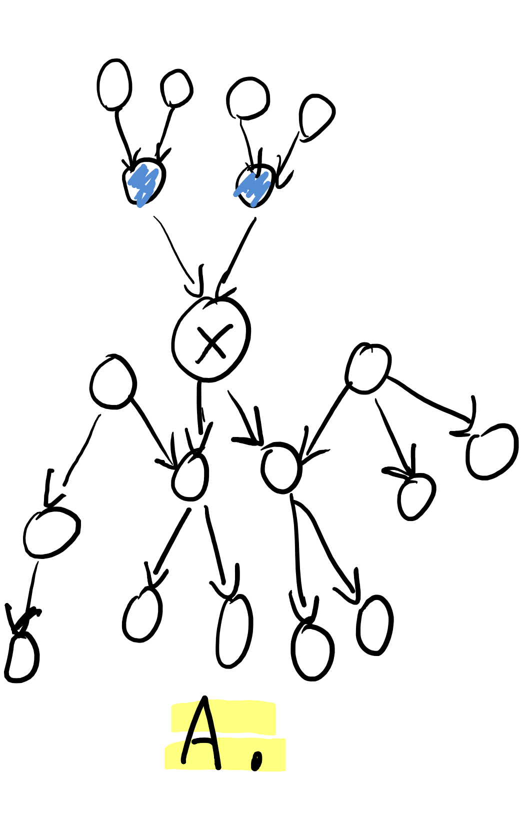
Independence (Markov Blanket)





What is the minimal set of nodes that must be observed in order to make **node X** independent from all the non-observed nodes in the network

Independence (Markov Blanket)



A node is conditionally independent from all the other nodes in the network, given its parents, children, and children's parents (i.e., its **Markov Blanket**) Configuration B

Variable elimination algorithm: Summary

$$P(Z, \cancel{Y_1 \ldots, Y_j} \quad \cancel{Z_1 \ldots, Z_j})$$

To compute $P(Z|\ Y_1{=}v_1, \ldots, Y_j{=}v_j)$:

1. Construct a factor for each conditional probability.
2. Set the observed variables to their observed values.
3. Given an elimination ordering, simplify/decompose sum of products
   - For all Z_i : Perform products and sum out Z_i
4. Multiply the remaining factors (all in ? $\boxed{Z}$)
5. Normalize: divide the resulting factor $f(Z)$ by $\sum_Z f(Z)$.

Variable elimination ordering

BNet with nodes {A B C D G}

$$P(G,D{=}t) = \boxed{\textstyle\sum_{A,B,C,}\, f(A,G)\, f(B,A)\, f(C,G,A)\, f(B,C)}$$

CBA

$$\textstyle\sum_A f(A,G) \sum_B f(B,A) \sum_C f(C,G,A)\, f(B,C)$$

BCA

$$\textstyle\sum_A f(A,G) \sum_C f(C,G,A) \sum_B f(B,C)\, f(B,A)$$

Complexity: Just Intuition…..

- **Tree-width of a network given an elimination ordering:** max number of variables in a factor created while running VE.
- **Tree-width of a belief network :** min tree-width over all elimination orderings (only on the graph structure and is a measure of the sparseness of the graph)

- **The complexity of VE is exponential in the tree-width** ☹ and linear in the number of variables.

- Also, finding the elimination ordering with minimum tree-width is NP-hard ☹ (but there are some good elimination ordering heuristics)

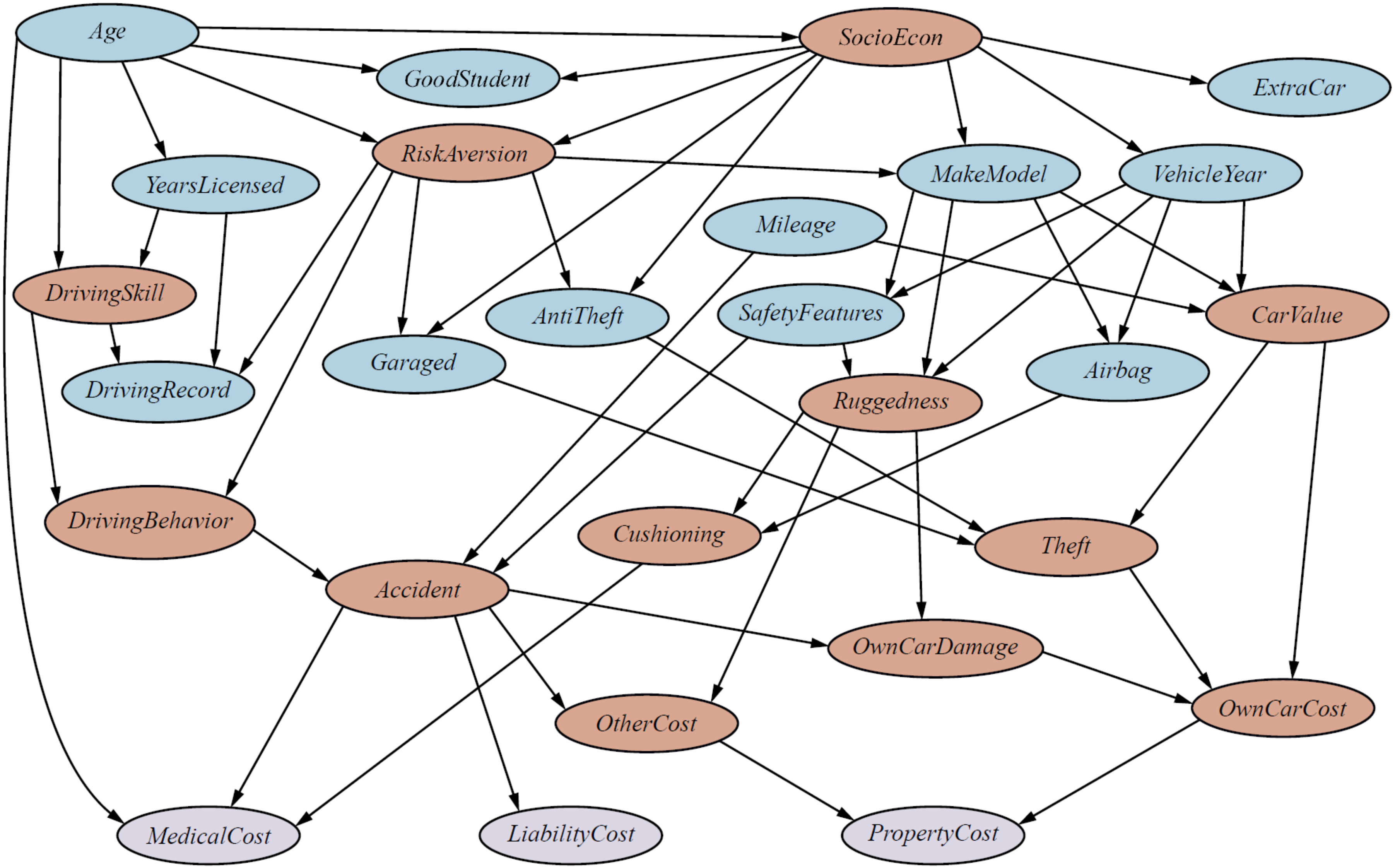


Figure 13.9 A Bayesian network for evaluating car insurance applications.

AIMA 4th Edition (Russell and Norvig) 2020

Lecture Overview

- **Recap of BNs** Representation and Exact Inference
- Start Belief Networks **Approx. Reasoning**
    - Intro to **Sampling**
    - First Naïve Approx. Method: **Forward Sampling**
    - Second Method: **Rejection Sampling**

Approximate Inference

Basic idea:

- Draw N samples from known prob. distributions
- Use those samples to estimate unknown prob. distributions

Why sample?

- Inference: getting a sample is faster than computing the right answer (e.g. with variable elimination)

We use *Sampling*

Sampling is a process to **obtain samples** adequate to estimate an **unknown probability**

How do we get samples?

Known prob. distribution(s)

Samples ←

↓

Estimates for unknown (hard to compute) distribution(s)

Generating Samples from a Known Distribution

For a random variable X with

- values $\{x_1,\ldots,x_k\}$
- Probability distribution $P(X) = \{P(x_1),\ldots,P(x_k)\}$

Partition the interval [0, 1] into k intervals p_i , one for each x_i , with length $P(x_i)$

To generate one sample

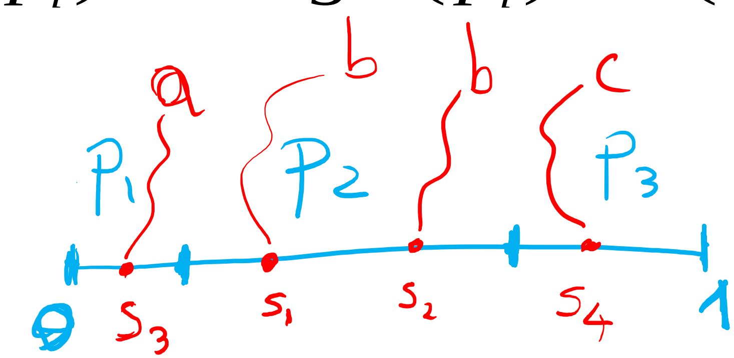
✓ Randomly generate a value y in [0, 1] (i.e. generate a value from a uniform distribution over [0, 1].

✓ Select the value of the sample based on the interval p_i that includes y

From probability theory: $P(y \subset p_i) = Length(p_i) = P(x_i)$

$X \{a, b, c\}$

X	P(x)
a	.1
b	.6
c	.3



From Samples to Probabilities

X	*count*
x_1	n_1
$\vdots$	$\vdots$
x_k	n_k
total	m

$\leftrightarrow$

X	*probability*
x_1	n_1/m
$\vdots$	$\vdots$
x_k	n_k/m

X	Count
00	~~2~~4342
10	258
01	301
11	2299
total	7200

e.g. $P(01) = \frac{301}{7200}$

Count total number of samples m

Count the number n_i of samples x_i

Generate the frequency of sample x_i as n_i / m

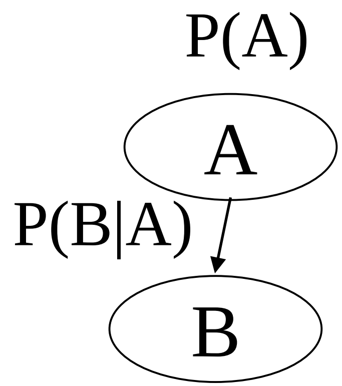
This frequency is your estimated probability of x_i

Sampling for Bayesian Networks (N)

- Suppose we have the following BN with two binary variables

A=1	0.3
A=0	0.7

A=1	B=1	0.7
	B=0	0.3
A-0	B=1	0.1
	B=0	0.9



Known distributions

- It corresponds to the joint probability distribution
  - $P(A,B) = P(B|A)P(A)$
- To sample from $P(A,B)$ i.e., unknown distribution
  - we first sample from $P(A)$. Suppose we get $A = 0$.
  - In this case, we then sample from…. $P(B|A=0)$
  - If we had sampled $A = 1$, then in the second step we would have sampled from $P(B|A=1)$

A=0 B=1
A=0 B=0
A=1 B=1

Prior (Forward) Sampling

\+

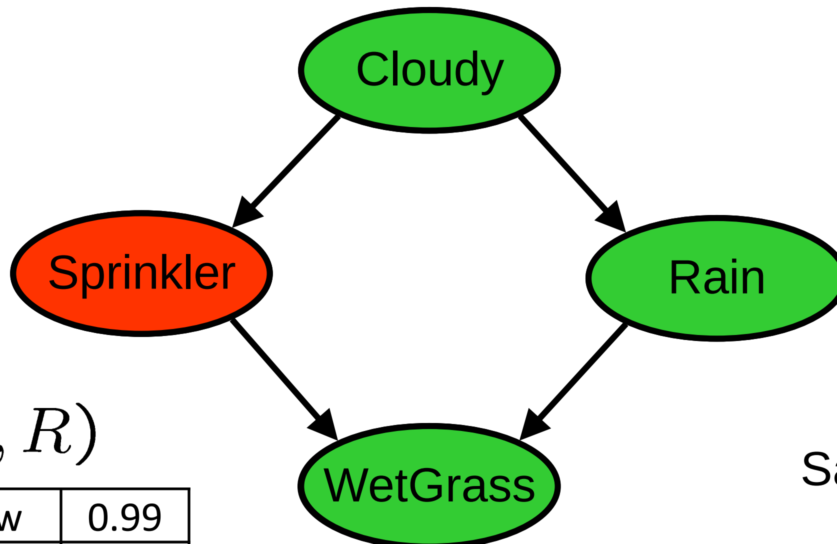
\-

$P(C)$

+c	0.5
-c	0.5

$P(S|C)$

+c	+s	0.1
	-s	0.9
-c	+s	0.5
	-s	0.5



$P(R|C)$

+c	+r	0.8
	-r	0.2
-c	+r	0.2
	-r	0.8

$P(W|S,R)$

+s	+r	+w	0.99
		-w	0.01
	-r	+w	0.90
		-w	0.10
-s	+r	+w	0.90
		-w	0.10
	-r	+w	0.01
		-w	0.99

Samples:

+c, -s, +r, +w

-c, +s, -r, +w

…

Example

We’ll get a bunch of samples from the BN:

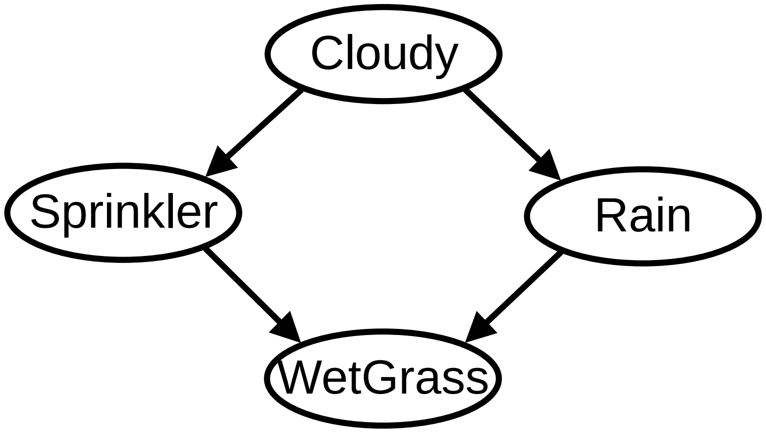
+c, -s, +r, +w

+c, +s, +r, +w

-c, +s, +r, -w

+c, -s, +r, +w

-c, -s, -r, +w



If we want to know P(W)

- We have counts <+w:4, -w:1>
- Normalize to get P(W) = <+w: .8 , -w: .2>
- This will get closer to the true distribution with more samples

Example

Can estimate anything else from the samples, besides P(W), P(R) , etc:

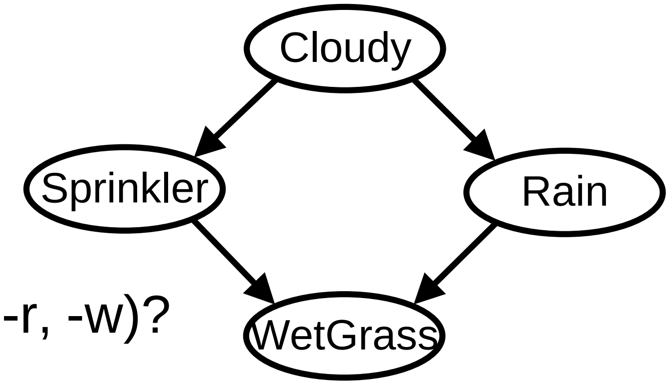
+c, -s, +r, +w

+c, +s, +r, +w

-c, +s, +r, -w

+c, -s, +r, +w

-c, -s, -r, +w



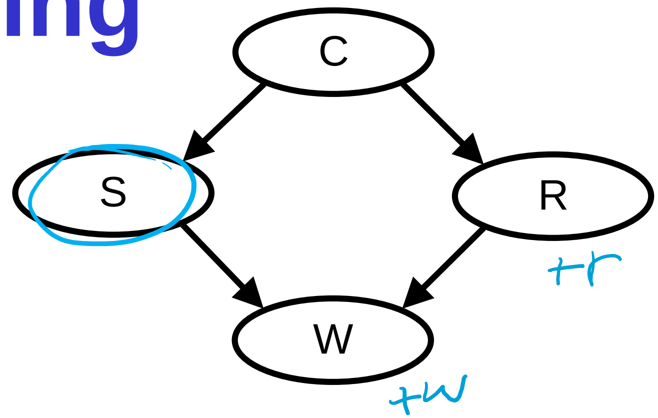
- What about P(C| +w)? P(C| +r, +w)? P(C| -r, -w)?

A. $\begin{array}{cc} +c & -c \\ \end{array} \begin{bmatrix} 0 & 1 \end{bmatrix}$ B. $\begin{array}{cc} +c & -c \\ \end{array} \begin{bmatrix} .5 & .5 \end{bmatrix}$ C. $\begin{array}{cc} +c & -c \\ \end{array} \begin{bmatrix} 1 & 0 \end{bmatrix}$

D. None of the above

Can use/generate fewer samples when we want to estimate a probability conditioned on evidence?

Rejection Sampling



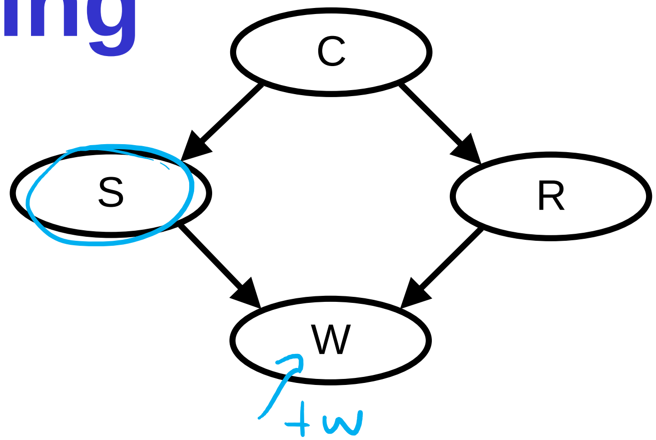
Let’s say we want P(S| +r, +w)

- Ignore (reject) samples which don’t have W=+w
- This is called rejection sampling
- It is also consistent for conditional probabilities (i.e., correct in the limit)

~~+c, -s, -r, +w~~
+c, +s, +r, +w
~~-c, +s, +r, -w~~
+c, -s, +r, +w
~~-c, -s, -r, +w~~

See any problem as the number of evidence vars increases?
Or the evidence is rare…

Rejection Sampling



Let’s say we want P(S| +w)

- Ignore (reject) samples which don’t have W=+w
- This is called rejection sampling
- It is also consistent for conditional probabilities (i.e., correct in the limit)

See any problem as the number of evidence vars increases?

+c, -s, +r, +w
+c, +s, +r, +w
-c, +s, +r, -w
+c, -s, +r, +w
-c, -s, -r, +w

References to applications to climate change and healthcare……..

Bnets to assess and manage Climate Change

Journal of Environmental Management

Volume 202, Part 1, 1 November 2017, Pages 320-331

Reviewing Bayesian Networks potentials for climate change impacts assessment and management: A multi-risk perspective

AnnaSperotto[a][b] José LuisMolina[c] SilviaTorresan[a][b] AndreaCritto[a][b] AntonioMarcomini[a][b]

One Recent Example from that review

Environmental Modelling & Software Journal

Volume 80, **June 2016**, Pages 132-142

A Bayesian Belief Network to assess rate of changes in coral reef ecosystems

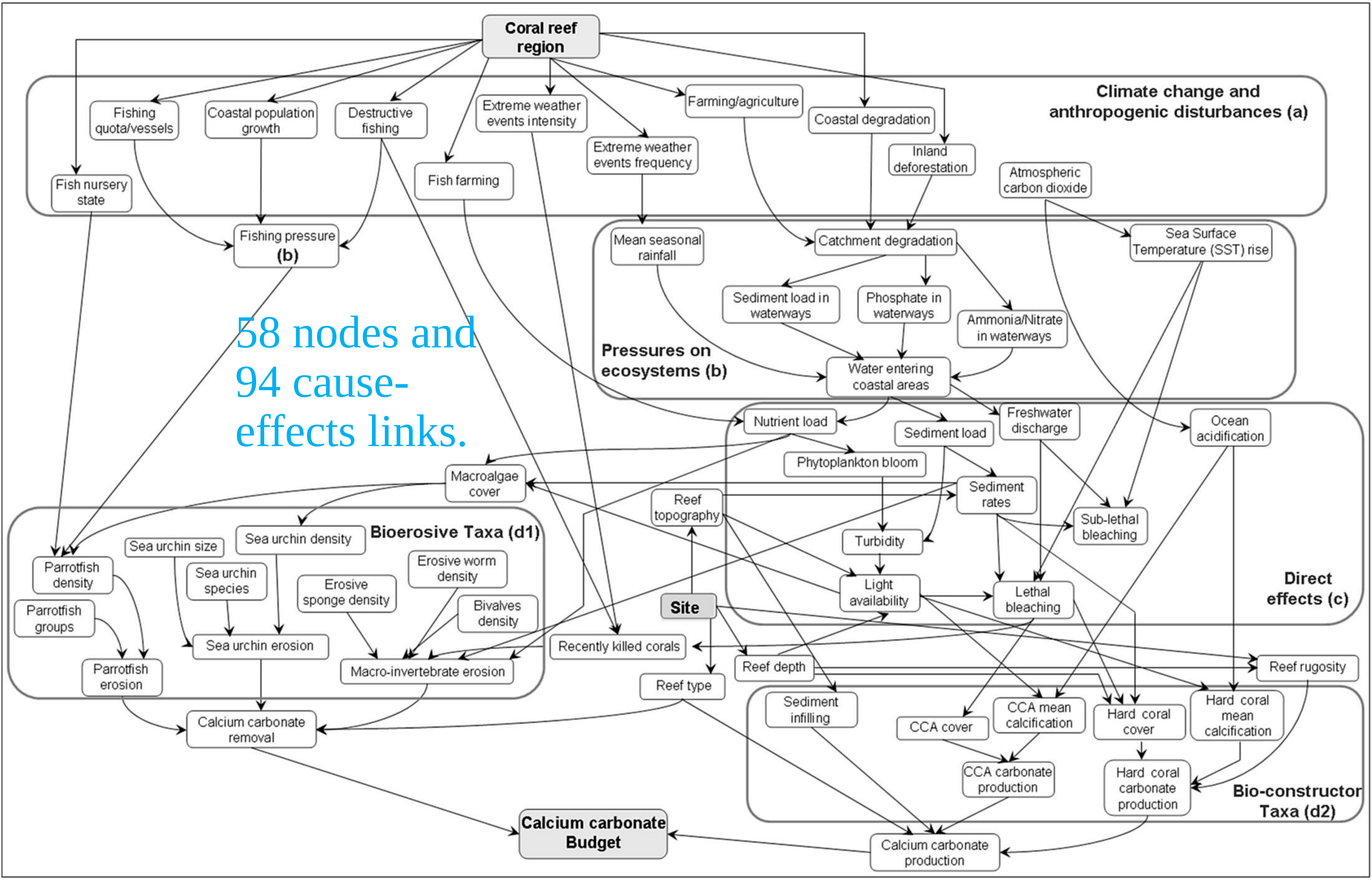
Coral Reef Research Unit, University of Essex, United Kingdom

St. George's University, Grenada

Department of Computer Science, Brunel University, United Kingdom

Carbonate Budget BBN (CARBNET)

- We propose a Bayesian Belief Network (BBN) approach, which offers a methodological framework to **address uncertainty** (Bennett et al., 2013, Kelly et al., 2013).
- Can **aid sustainable coral reef management** and prevent further decline.
- Help **evaluate effects of anthropogenic and climatic disturbances** on the reef framework
- Consider **impacts of implementing management interventions** or decision options in order to **maximize their benefit** (Uusitalo et al., 2015).
- **CARBNET**: developed to evaluate coral reef $CaCO_3$ (carbonate) balance under changing environmental conditions and across reef bioregions.



58 nodes and 94 cause-effects links.

CARBNET Engineering

- Variables identified through **literature search**
- Nodes representing different levels of spatial resolution were used to capture changes that may occur at different spatial scales.
- Presence/absence of **reef-building and erosive organisms or reef growth and erosion processes** are captured at the smallest scale of reef depth, but also for an entire reef (‘*Site*’), sub-region (‘*Reef type*’, ‘*Reef topography*’) or region (‘*Coral reef region*’).
- The CARBNET conceptualisation was proposed to **twenty experts in the field of coral reef management and ecology** to identify flaws in the network structure and address structural bias before model parameterisation.

Another Example

178 *Water Quality: Current Trends and Expected Climate Change Impacts* (Proceedings of symposium H04 held during IUGG2011 in Melbourne, Australia, July 2011) (IAHS Publ. 348, 2011).

Predicting water quality responses to a changing climate: building an integrated modelling framework

F. DYER[1], S. EL SAWAH[2], E. HARRISON[1], S. BROAD[1], B. CROKE[2], R. NORRIS[1] & A. JAKEMAN[2]

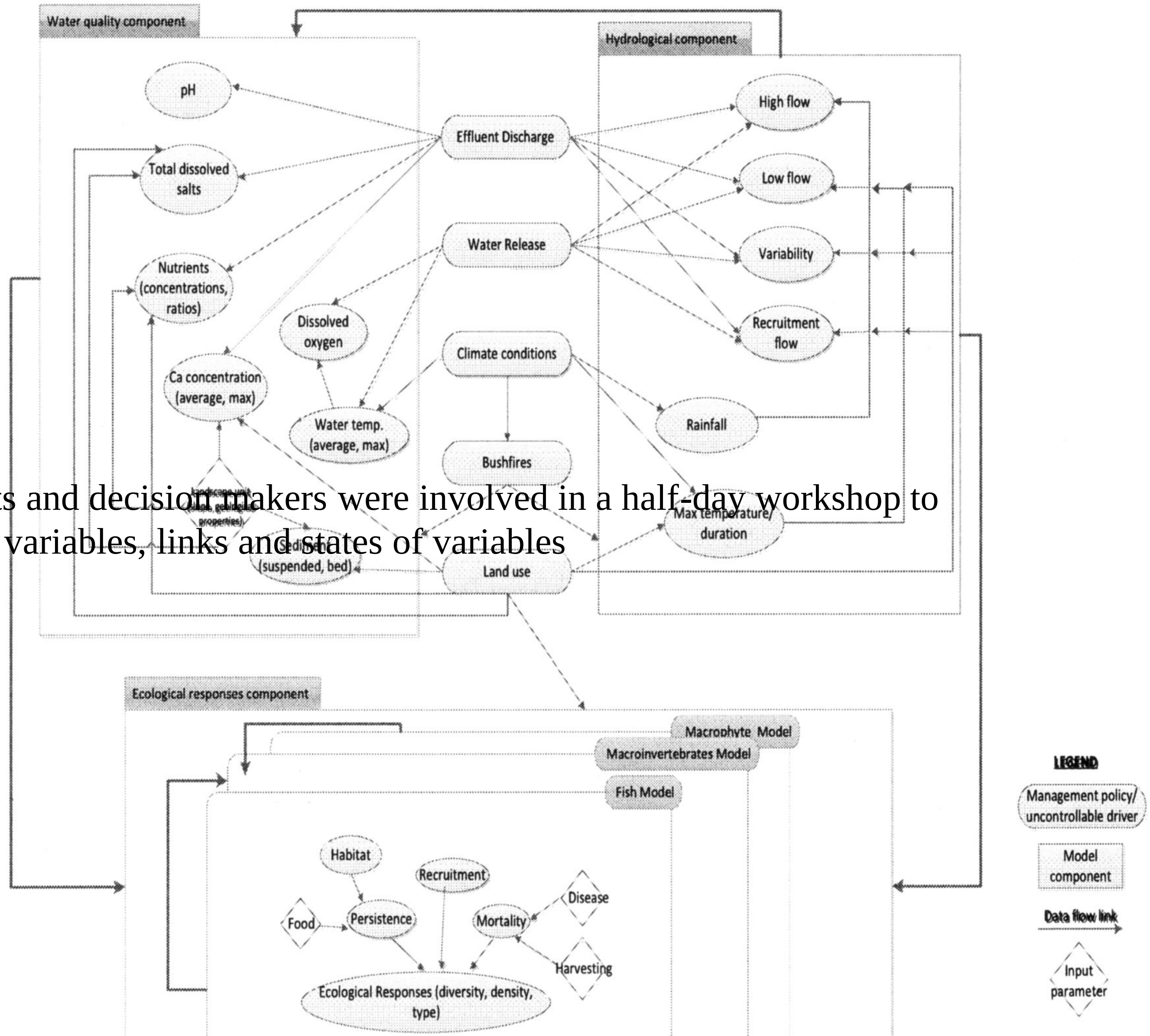
1 *Institute for Applied Ecology, University of Canberra, Canberra, Australia*
fiona.dyer@canberra.edu.au

2 *Integrated Catchment Assessment and Management Centre, National Center for Groundwater Research and Training, Australian National University, Canberra, Australia*

Abstract The future management of freshwater resources for human and environmental needs requires an integrated set of tools for predicting the relationship between climate change, water quality and ecological responses. In this paper, we present the early phases of a project for building a Bayesian network (BN) based framework to link ecological and water quality responses to features of the flow regime in the Molonglo and Yass rivers in southeastern Australia. At this stage, the objective is to conceptualize the modelling components and define causal links. Expert elicitation was used to identify important drivers and interactions which influence water quality attributes and related ecological responses.

Key words Bayesian network models; water quality; prediction; climate change; integrated modelling

Corresponding BNet



group of 14 experts and decision makers were involved in a half-day workshop to
fine the important variables, links and states of variables

Many applications in Health Care

McLachlan S, Dube K, Hitman GA, Fenton NE, Kyrimi E (2020) . **Bayesian networks in healthcare: Distribution by medical condition .** *Artificial Intelligence in Medicine* vol. 107 , Article 101912 , 101912 - 101912 .**10.1016/j.artmed.2020.101912**

https://qmro.qmul.ac.uk/xmlui/handle/123456789/65190

Learning Goals for today's class

➢ You can:

- Motivate the need for approx. inference in Bnets
- Describe and compare Sampling from a single random variable
- Describe and Apply Forward Sampling in BN
- Describe and Apply Rejection Sampling

TODO for Fri

- Read textbook
  - 8.6.3 Rejection Sampling
  - 8.6.4 Likelihood Weighting
- Assignment-2 will be out tonight…
- Next research paper will be **this coming Mon**

Hoeffding's inequality

- Suppose *p* is the true probability and *s* is the sample average from *n* independent samples.

$$P(|s-p|>\varepsilon)\le 2e^{-2n\varepsilon^2}$$

< δ

- *p* above can be the probability of any event for random variable *X* = $\{X_1,\ldots X_n\}$ described by a Bayesian network

- If you want an infinitely small probability of having an error greater than *ε*, you need infinitely many samples
- But if you settle on something less than infinitely small, let's say δ, then you just need to set

$$2e^{-2n\varepsilon^2}<\delta$$

- So you pick
  - the error *ε* you can tolerate,
  - the frequency *δ* with which you can tolerate it
- And solve for *n*, i.e., the number of samples that can ensure this performance

$$n>\frac{-\ln\frac{\delta}{2}}{2\varepsilon^2} \qquad (1)$$

Hoeffding's inequality

- Examples:
  - You can tolerate an error greater than 0.1 only in 5% of your cases
  - Set ε =0.1, δ = 0.05
  - Equation (1) gives you n > 184

$$n > \frac{-\ln \frac{\delta}{2}}{2\varepsilon^2} \qquad (1)$$

can rewrite as

$$n > \frac{\ln \frac{2}{\delta}}{2\varepsilon^2}$$

- If you can tolerate the same error (0.1) only in 1% of the cases, then you need 265 samples

- If you want an error greater than 0.01 in no more than 5% of the cases, you need 18,445 samples

so it should be clear that

$\varepsilon \downarrow$

$\delta \downarrow$

$n \uparrow$

$\downarrow$ goes down

$\uparrow$ goes up