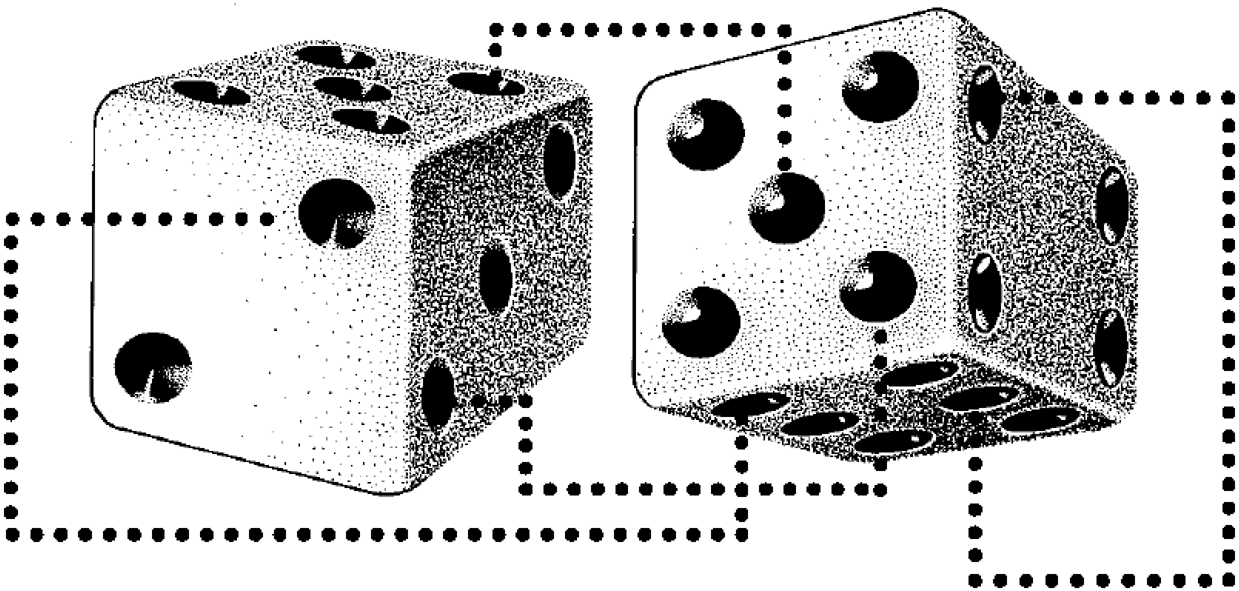
Bayesian Networks



CMEE

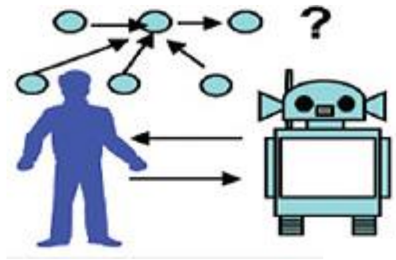
Véronique Lefebvre

v.lefebvre@imperial.ac.uk

Imperial College London

Bayesian Networks overview

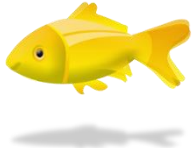
- What are they?
  - Web of quantified interactions between random variables
- What are they for?
  - Reasoning under uncertainty from experience / management of trade-offs
  - Stage between a statistical model and a mechanistic model
    - where our understanding of processes/associations is incomplete
    - And there is an ongoing data collection and gain of expertise



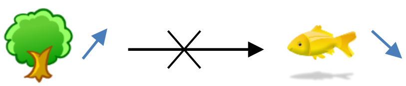
- What do they do?
  - Aggregate expert / statistical knowledge and make it explicit
  - Give global insight based on local observations
  - Predict what will happen given a combination of causes
  - or infer causes from a combination of observed effects
  - Update and learn as more observations are incorporated

A toy example: the “Native Fish” Bayesian network

Consider a population of river fish that may be at risk :

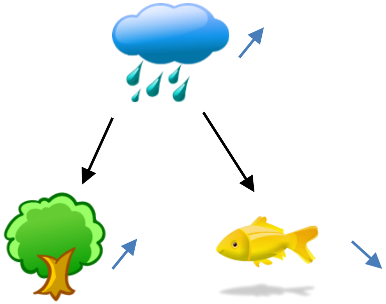


A study discovers a strong statistical inverse correlation between the fish abundance and the condition of the trees lining the river:



However in an area where trees have been cut the abundance of fish was not affected, so the trees do not causally affect the fish.

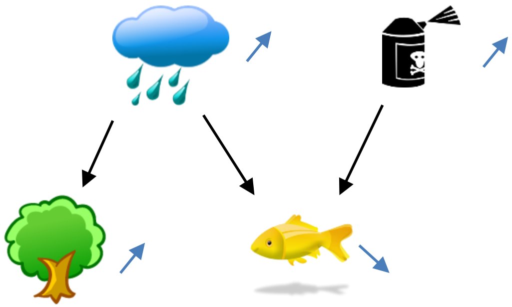
The variable rainfall is correlated with both, when rainfall changes both trees and fish are affected. Rainfall is considered causally responsible for the other two.

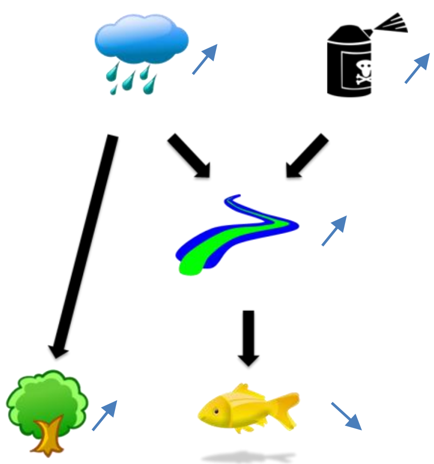


Rainfall directly affects the trees condition, however the study needs to explicit the relation between rainfall and fish…..

A toy example: the "Native Fish" Bayesian network

The study now includes the use of pesticide in croplands next to the river. The correlation between rainfall and fish abundance is weakened; but rainfall and pesticide used together are able to predict the fish abundance.

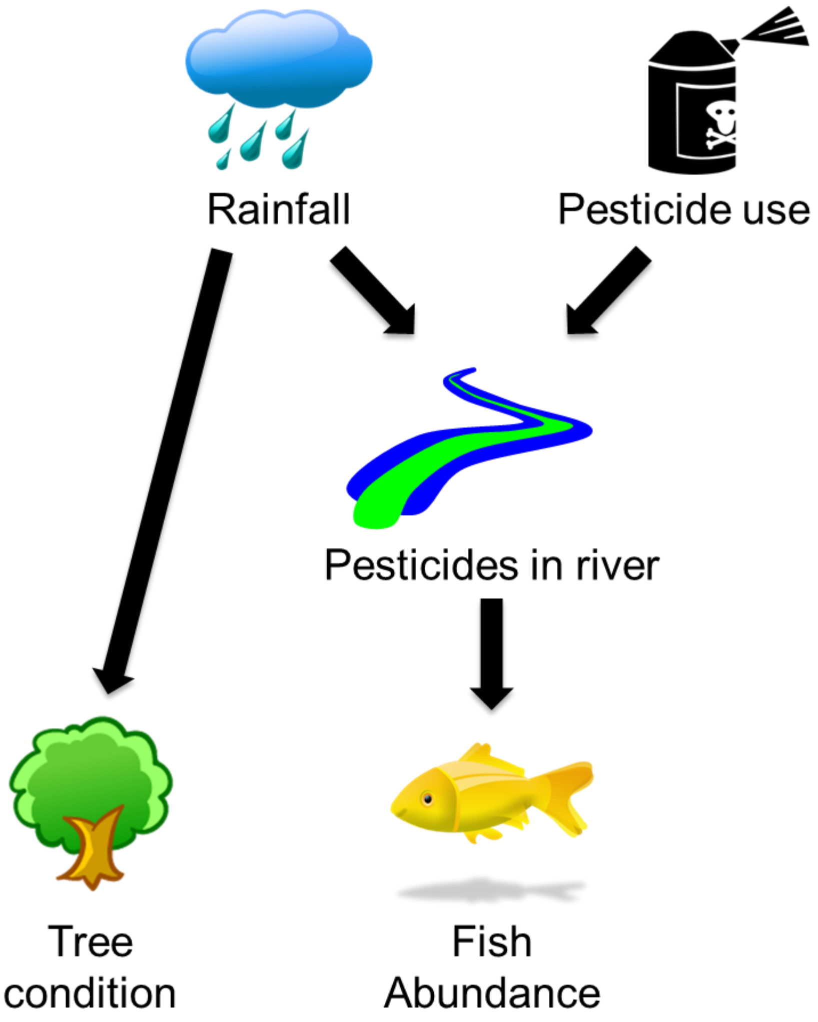




The rain washes the pesticides used on crops into the river. The variable concentration of pesticides is included, it is found to be well predicted by rainfall and pesticide, and is a good predictor for fish abundance.

A toy example: the “Native Fish” Bayesian network

- Nodes are key random variables
- Arcs between nodes show assumptions of dependence, here interpreted as causality
- But causal connections are not absolute
  - E.g. sometimes it will rain but the tree condition will not improve much, because of a severe drought for example
- Missing arcs represent assumptions of independence
  - E.g. pesticide use is unrelated to rainfall, trees are not affected by pesticides
  - No other causes of fish abundance is as important as the concentration of pesticides in the river
- Indirect arcs represent conditional independence
  - E.g. If I know the concentration of pesticides in the river, I do not need to investigate the pesticides use or the rainfall to predict the fish abundance

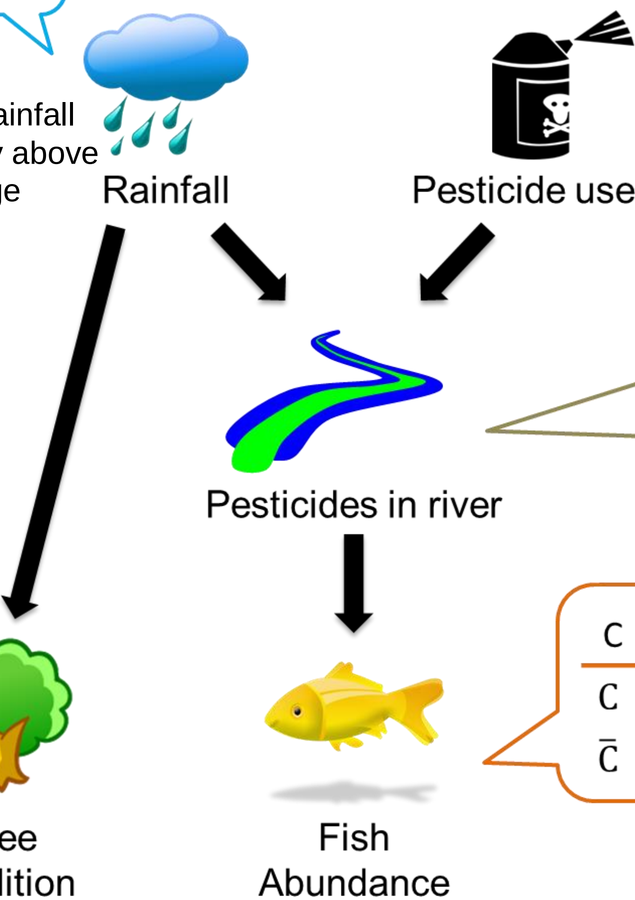


Adapted from A. Nicholson, O. Woodberry and C. Twardy (2010). The "Native Fish" Bayesian networks.

Toy example – quantifying dependence relations

- The structure represents local dependence, e.g. C= f(R, U)
- Probability distributions are used to represent uncertainty,
- prior distributions for root nodes,
- conditional distributions when the node has parents,
- Collectively referred to as Conditional Probability Tables (CPT)

Variables are specified as binary here, with states: 'high' and 'not high'



P(R)
0.55

e.g. Rainfall slightly above average

P(U)
0.7

Pesticides are used in 70 % of croplands around the river (or I believe there is a 70% chance of farmers using them)

R	U	P(C|R,U)
R	U	0.9
R	$\bar{U}$	0.4
$\bar{R}$	U	0.3
$\bar{R}$	$\bar{U}$	0.05

With high pesticide use, 90% risk of high concentration in river if high rainfall, only 30% if low rainfall

R	P(T|R)
R	0.8
$\bar{R}$	0.5

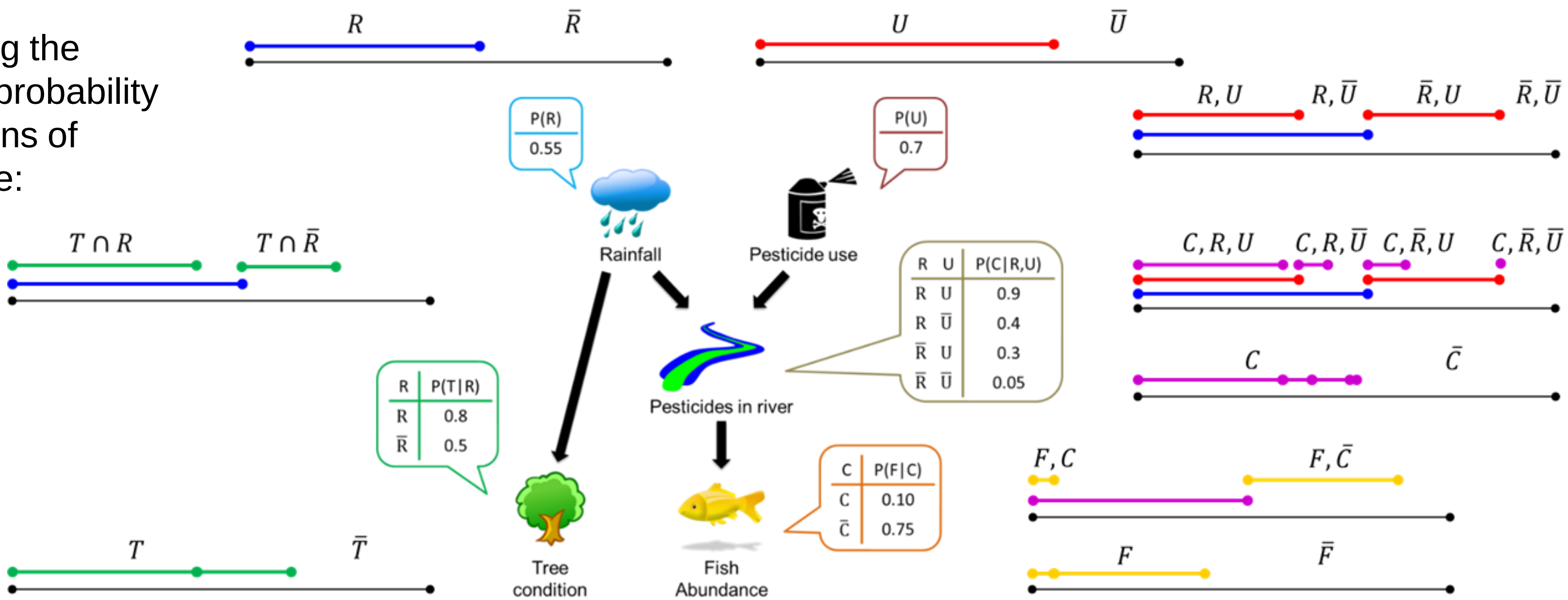
With high rainfalls 80% of trees are in good condition, but there are only 50 % when rainfall is low

C	P(F|C)
C	0.10
$\bar{C}$	0.75

75% chance of high fish abundance when the concentration of pesticides in the river is low

Toy example – BN evaluation

Computing the marginal probability distributions of each node:



e.g.:

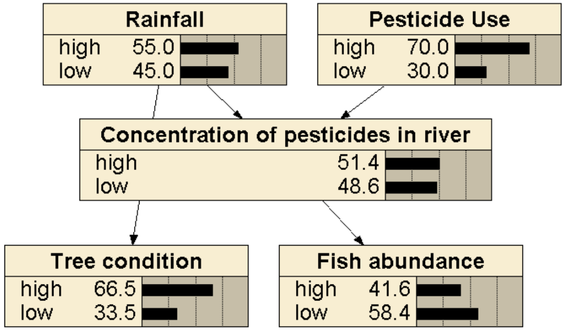
P(T)=P(T|R)P(R) + P(T|¬R)P(¬R)

P(T)=0.8x0.55 + 0.5x0.45 = 0.665

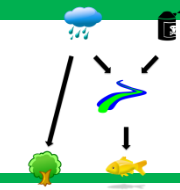
P(T) <=> P(T='high')

Based on past studies and expertise, the probability for the fish abundance to be high there is less than chance, 42 %.

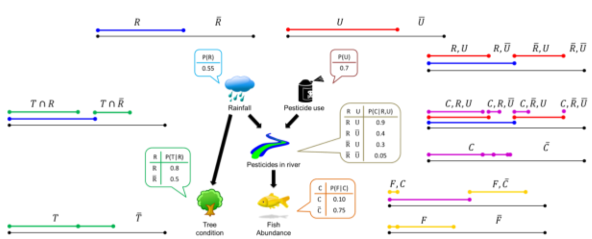
BN evaluation without observation:

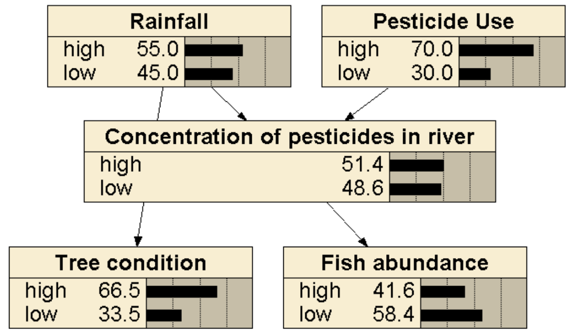


Toy example – Inference and Reasoning

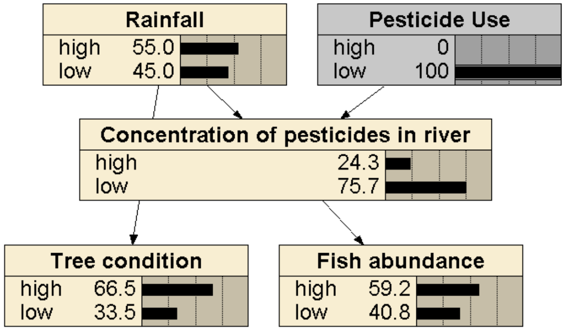


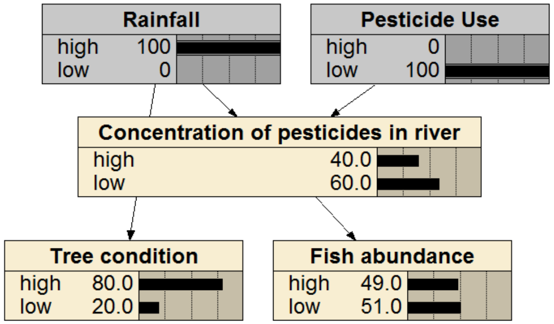
Evaluating the BN in different scenarios: probabilities computed given the evidence

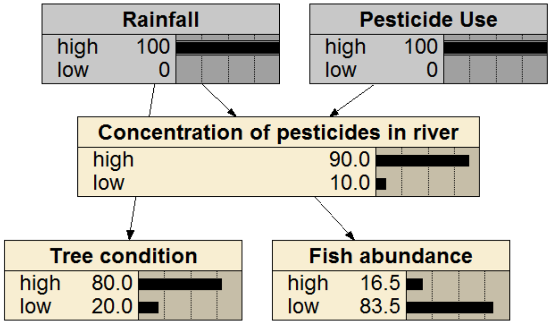




Predictive reasoning

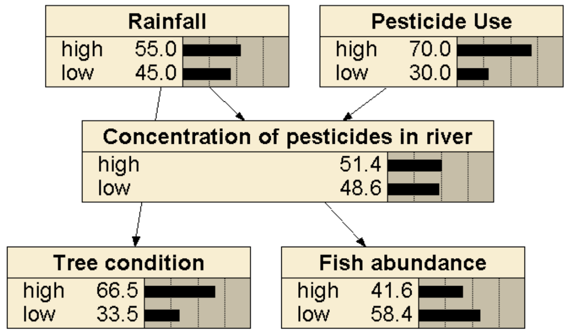




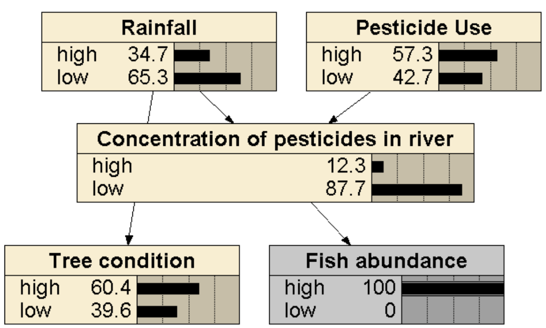


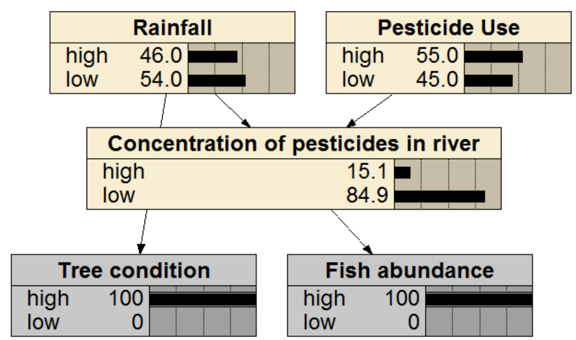
Toy example – Inference and Reasoning

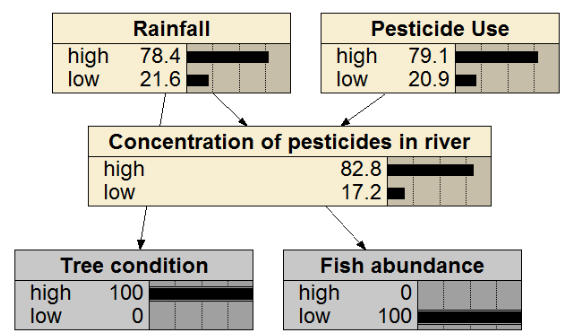
Evaluating the BN in different scenarios: probabilities computed given the evidence



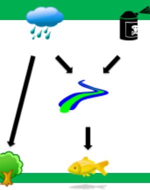
Diagnostic reasoning



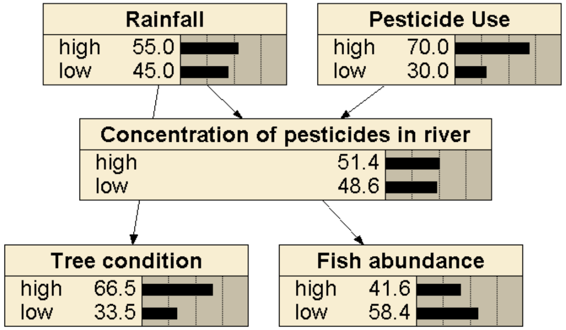




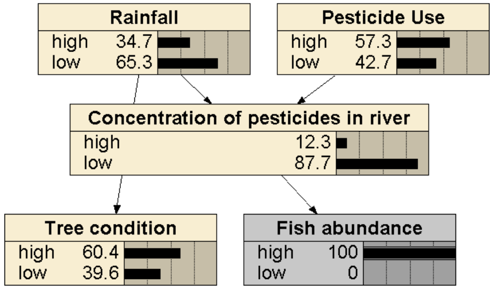
Toy example – Inference and Reasoning

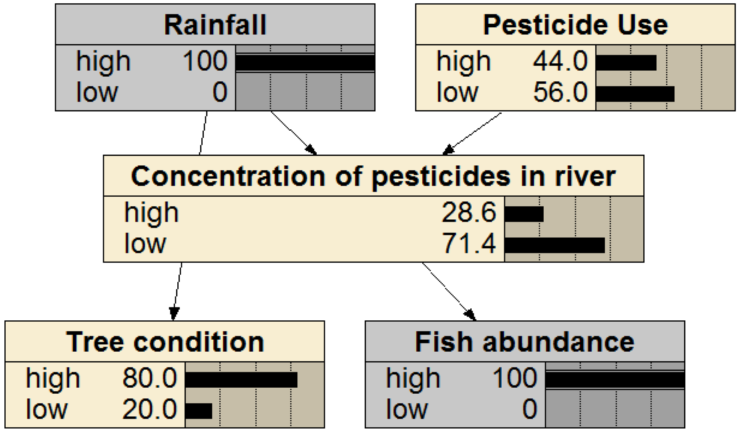


Evaluating the BN in different scenarios: probabilities computed given the evidence



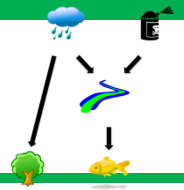
Combined reasoning



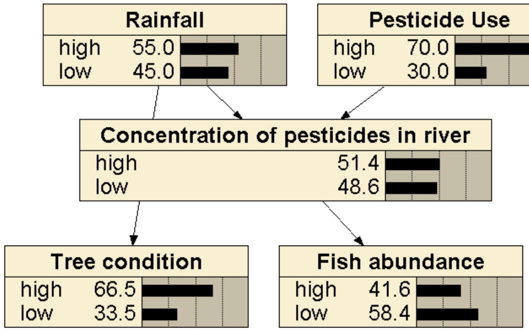


E.g. We want to infer the pesticide use for a high Fish abundance, under the worst case scenario of high rainfall:

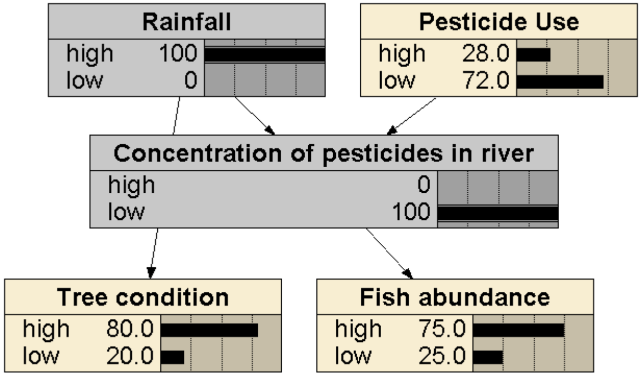
Toy example – Inference and Reasoning



Evaluating the BN in different scenarios: probabilities computed given the evidence



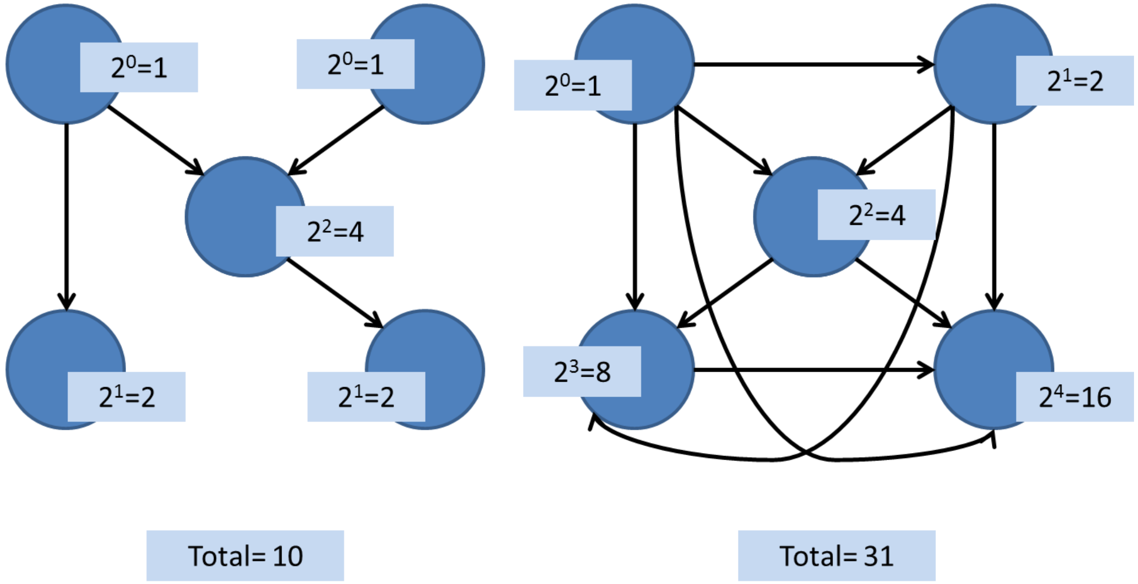
Inter-causal reasoning



Rainfall and pesticide use are independent, unless we know the concentration of pesticides in the river, then evidence of the rainfall state refines the diagnostic on pesticide use (but does not affect the prediction of fish abundance).

BN evaluation: Compactness

- Not all variables are linked: some are independent or conditionally independent (indirectly linked)
  - E.g U and R are independent, T and F are conditionally independent given R, F and R are conditionally independent given C
- Not all combinations of conditional probabilities need to be calculated
- The number of probabilities for each binary node is 2 to the power of the number of parents

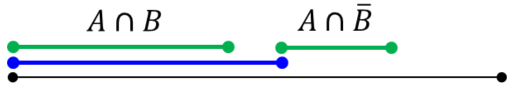


- *Key idea:* exploit regularities and structure of the domain, and **direct** causation to build the graph
- Conditional independence in BNs make them powerful tools to model complex situations with uncertainty (but not as simplified as Naïve Bayes)
- The BN evaluates all combinations from a small set of probabilities (compact representation)

Probabilities: Bayesian vs Frequentists

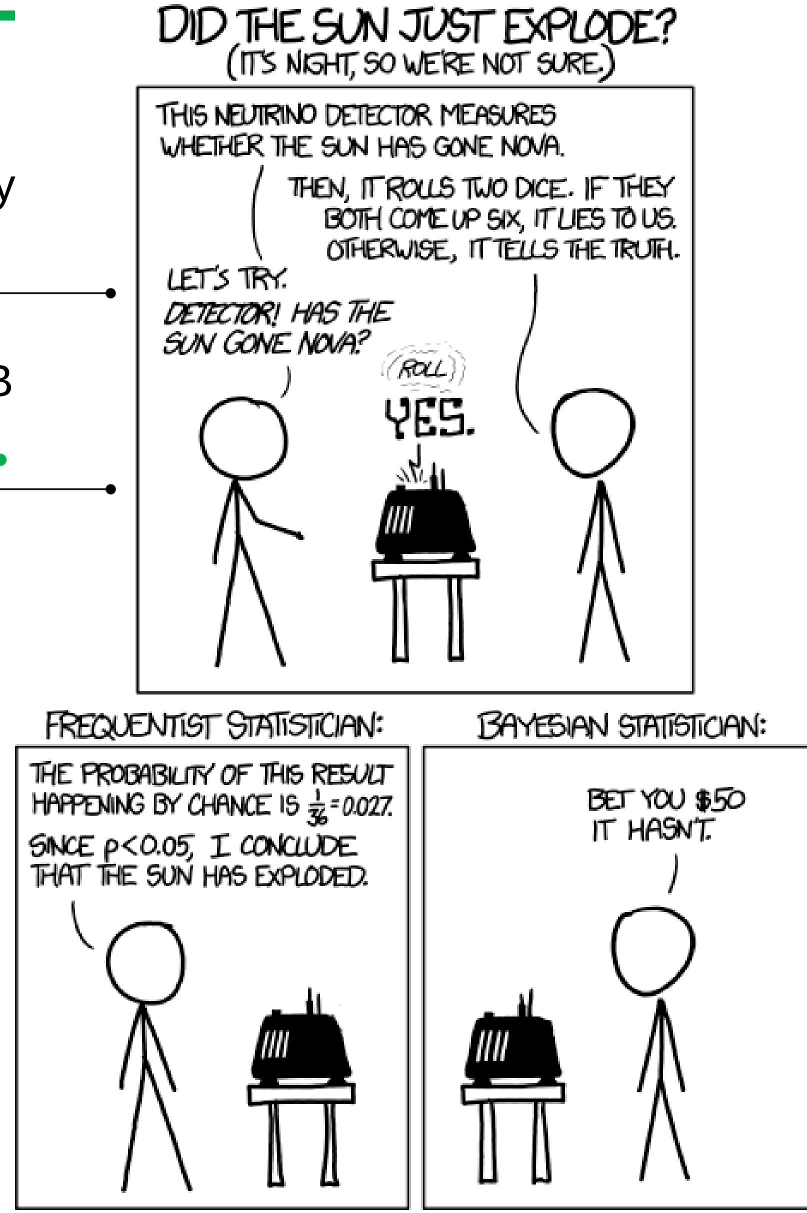
- Frequentist
  - An experiment is performed many times
  - P(A): Proportion of outcomes with property A out of all possible outcome

  - P(A|B): Proportions of outcomes with property A out of outcomes with property B



- Bayesian
  - Proposition A
  - P(A), prior, initial degree of belief in A
  - P(A|B), posterior, revised degree of belief having accounted for B

- Frequentist statistics rely on hypothetical repeated trials,
- Bayesian statistics involves “beliefs” but conditioned on observed evidence (expertise).



http://xkcd.com/

Bayesian networks can be … not Bayesians

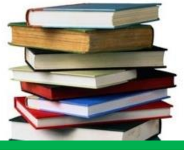
- Bayesian networks do not necessarily imply a commitment to Bayesian statistics (nor does Bayes’ Theorem)
- Frequentists methods are commonly used to estimate the priors and conditional probabilities

- BNs are so called because they use Bayes' Theorem for probabilistic inference

- But they can include Bayesian statistics (e.g. to represent parameters as random variables)

- Advantages in “dining at both tables”

Bayes' Theorem



- Bayes theorem can have a Frequentist or Bayesian interpretation:
Outcome given sample space, or revised initial belief

- It is used in BNs to propagate observation information
(from root to leaf: prediction, from leaf to root: inference)
- From the chain rule:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

- Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

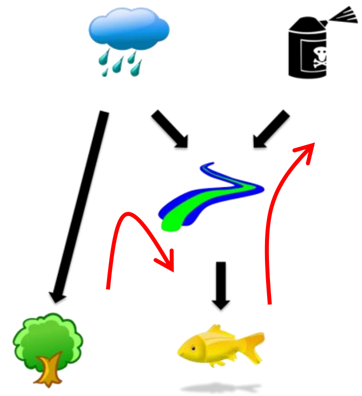
posterior probabilities are proportional to the numerator: likelihood times prior.

In the case of a binary partition:

$$P(B) = P(B \cap A) + P(B \cap \bar{A})$$

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$



The probability of interest may be P(A|B) or P(B|A)
depending on the node queried (cause or outcome)

Bayes' Theorem: Question

Consider a population where:

- 80% of lung cancer patients are smokers
- 10% of people have lung cancer
- 50% of people smoke

- What is the risk of a smoker to have lung cancer?

(or what % of smokers are lung cancer patients?)

0 – 25%? 25 – 50%? 50 – 75%? 75 – 100%?

Smoking kills

Bayes’ Theorem: Answer

- Answer:
  - 16%

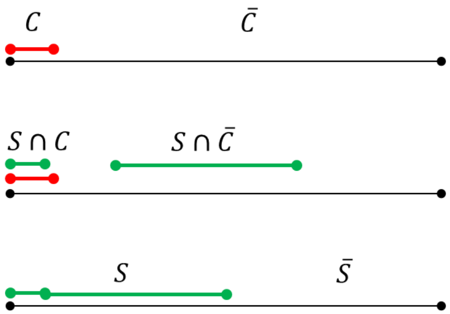
  - Probability that a lung cancer patient is a smoker: $P(S|C)=0.8$
  - Probability to have lung cancer: $P(C)=0.1$
  - Probability to smoke: $P(S)=0.5$

  - Probability that a smoker has cancer:
  - $P(C|S) = P(S|C) \times P(C) / P(S) = 0.8 \times 0.1 / 0.5 = 0.16$ (8 smokers out of 50)

  - The probability of someone having cancer, given that they are a smoker increases from 10% to 16%. Not a dramatic increase. But the probability to have cancer given non smoker is 4%.

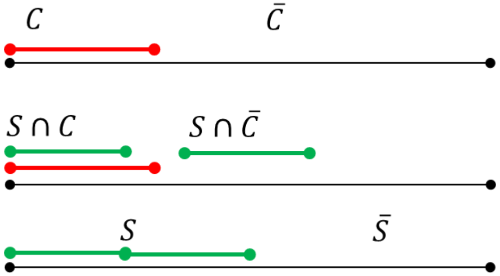
  - Why?
    - High proportion of smokers in cancer population, but ‘low’ overall cancer population compared to smokers population
    - The probability that a person that do not have cancer is a smoker is 46%:

      $P(S|\neg C) = (P(S) - P(S|C)P(C)) / P(\neg C) = 42/90$

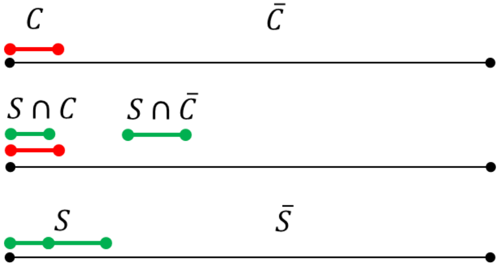


Bigger proportion of smokers that have cancer where:

Cancer proportion is larger:

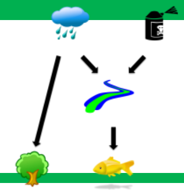


Smokers proportion is smaller:



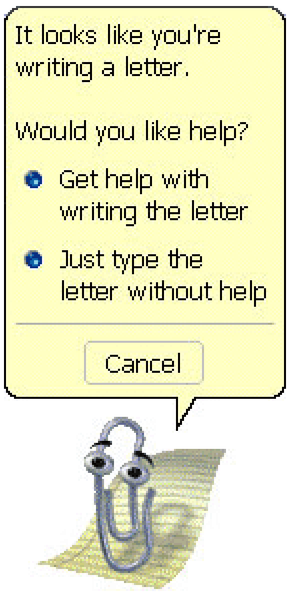
More at http://yudkowsky.net/rational/bayes

BN: Applications



- AI (speech recognition, machine translation, page ranking, robot path finder, behaviour prediction of phone/robot users, synthetic characters, recommendation systems)

- Medical diagnosis (to infer likelihood of different causes given observed symptoms)
- Industrial fault diagnosis / risk assessment
- Weather prediction
- Credit score
- Genetics (Gene regulatory networks)
- Aggregating expert forecasts
- Data mining
- Pattern recognition
- Resources management / Strategic planning
- decision support (e.g. conservation management)
- And Ecology



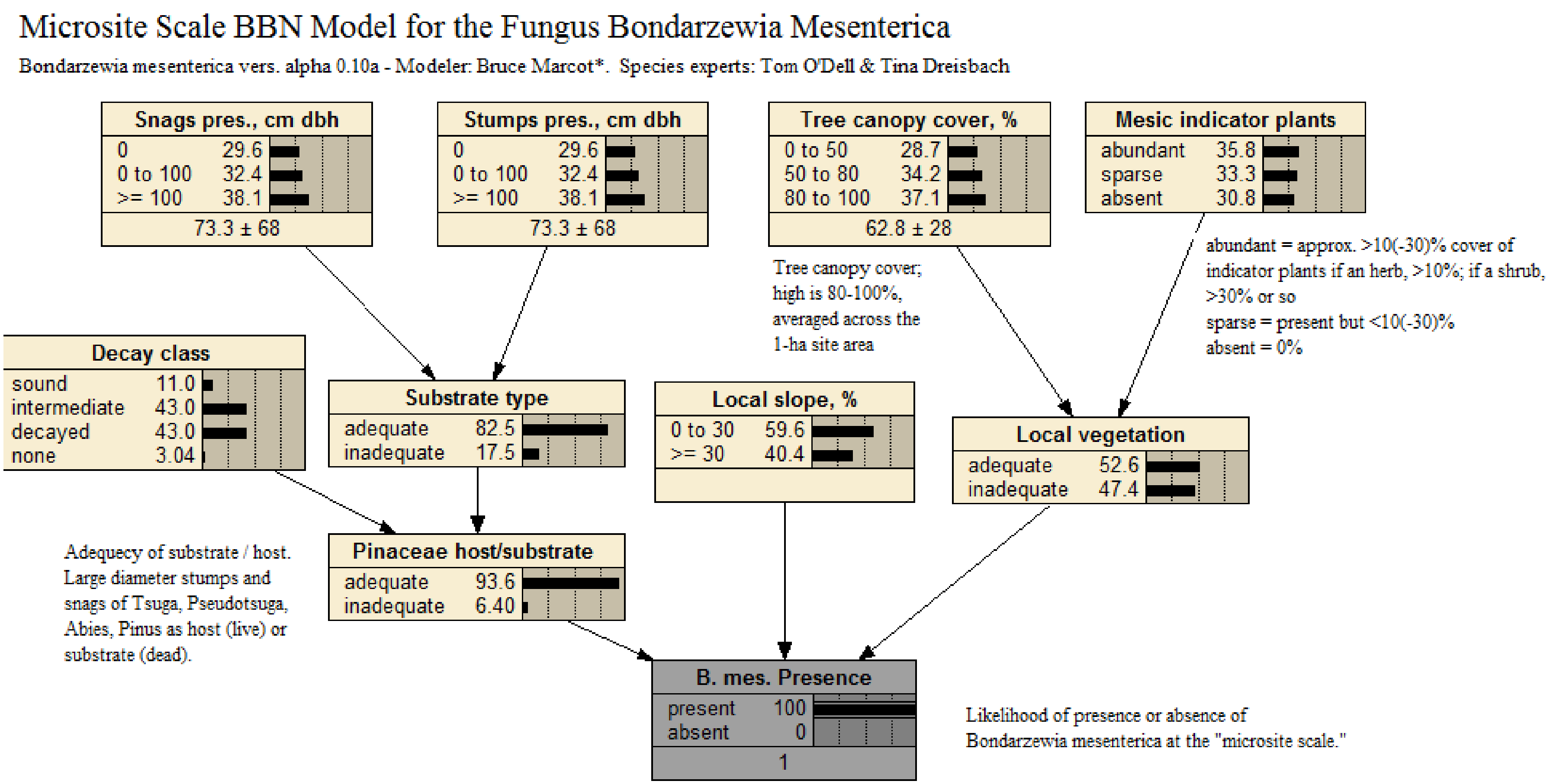
- BNs are found everywhere there are unknown processes but known dependencies
- And where data keep coming in

BN: Applications in Ecology

- To infer biological network structures from data :
  - relationship between ecological variables
    - (e.g., species' occurrence, species traits, communities, or ecosystems, and their functional behaviour)
  - and sample measurements obtained from a set of possibly related observations
- BNs are used in conservation and wildlife management
  - to depict the influence of habitat or environmental predictor variables on ecological-response variables.
    - BNs predict the probability of ecological responses to varying input assumptions such as habitat and population-demography conditions.
    - BNs serve well as part of a risk-management framework by explicitly displaying the "causal web" of interacting factors and the probabilities of multiple states of predictor and response variables.

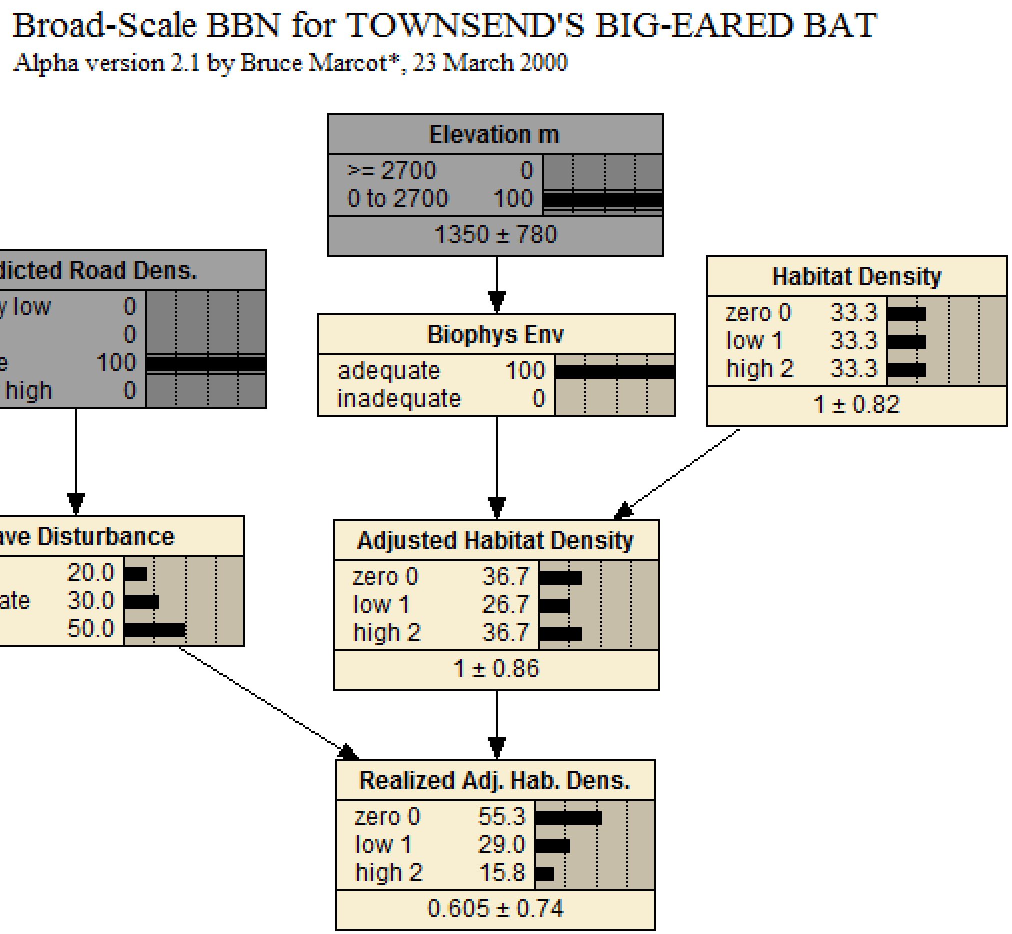
Examples in the Netica (Norsys) Bayes nets library:
http://www.norsys.com/netlibrary/index.htm

Presence of Fungus B



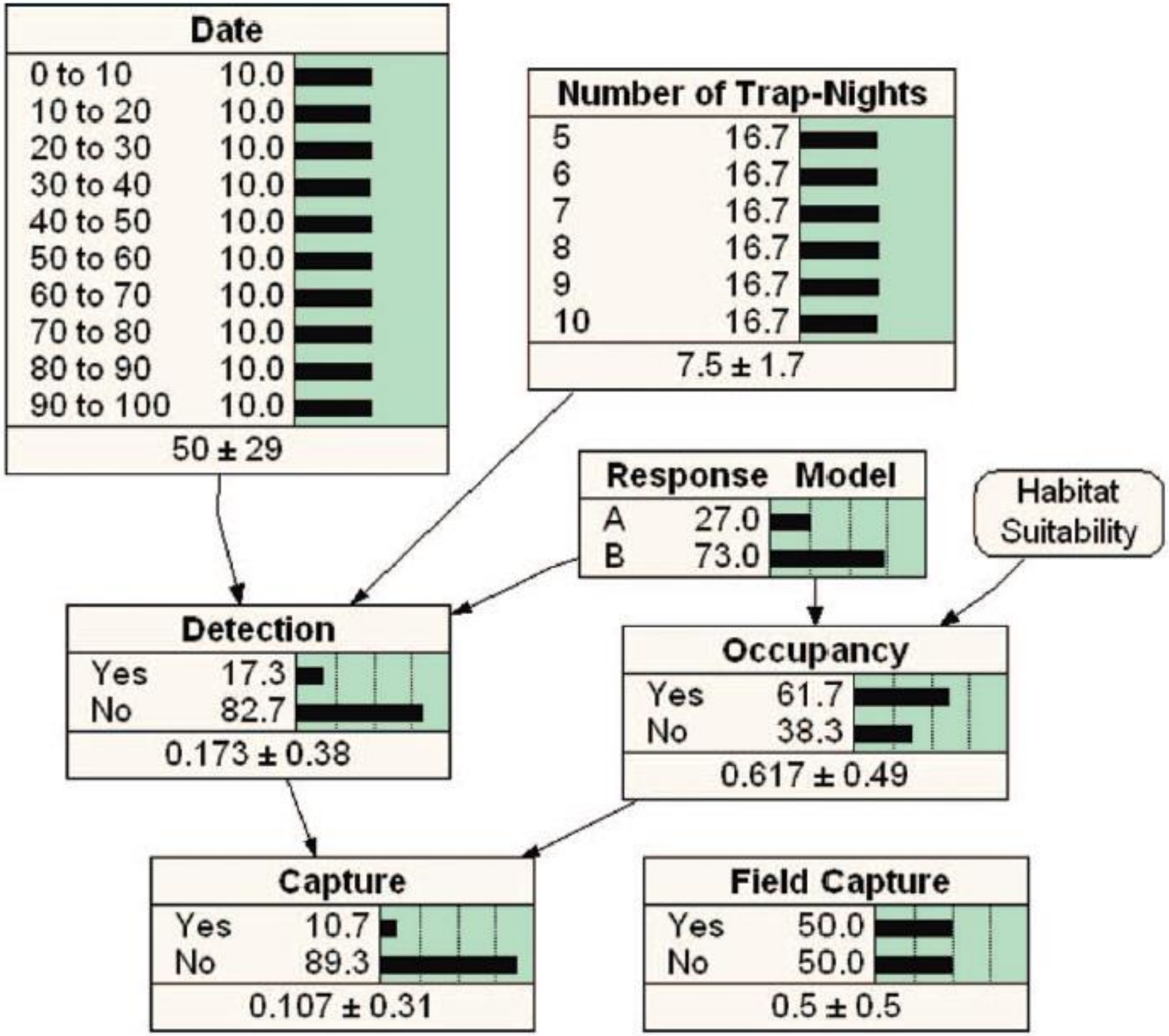
Marcot, B. G. 2006. Characterizing species at risk I: modeling rare species under the Northwest Forest Plan. *Ecology and Society* **11**(2): 10. [online] URL: http://www.ecologyandsociety.org/vol11/iss2/art10/

Bat habitat density



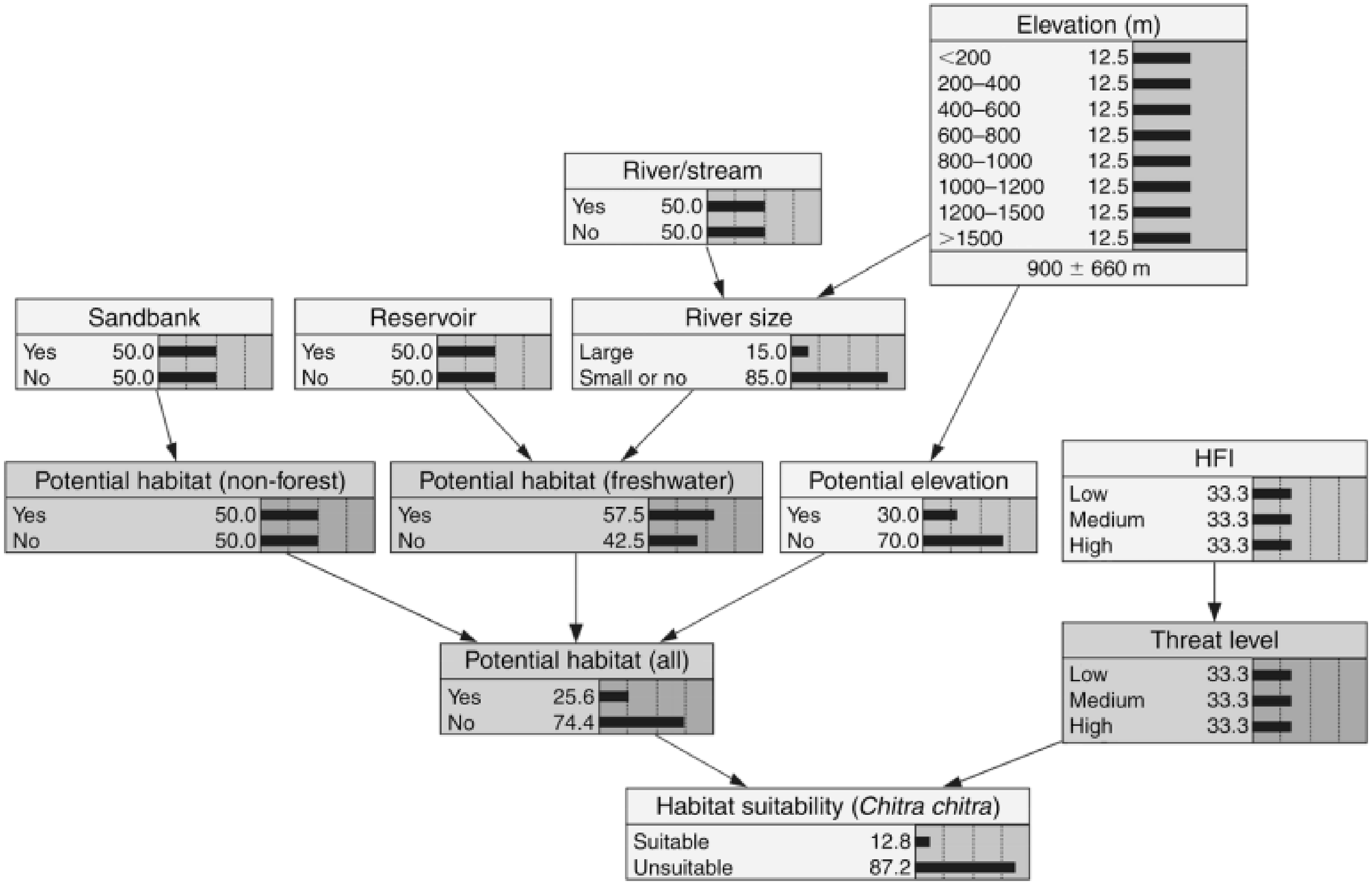
Marcot, Bruce G., et al. "Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement." *Forest ecology and management* 153.1 (2001): 29-42.

Factors affecting expected capture of northern flying squirrels

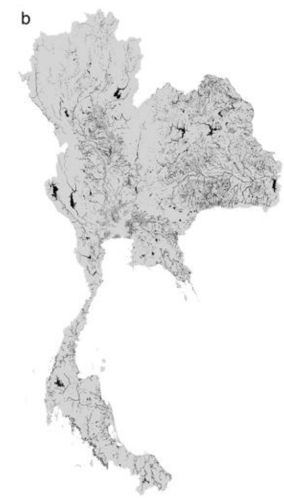


Marcot, Bruce G., et al. "Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation." *Canadian Journal of Forest Research* 36.12 (2006): 3063-3074.

Habitat suitability modelling (here for a species of turtle)



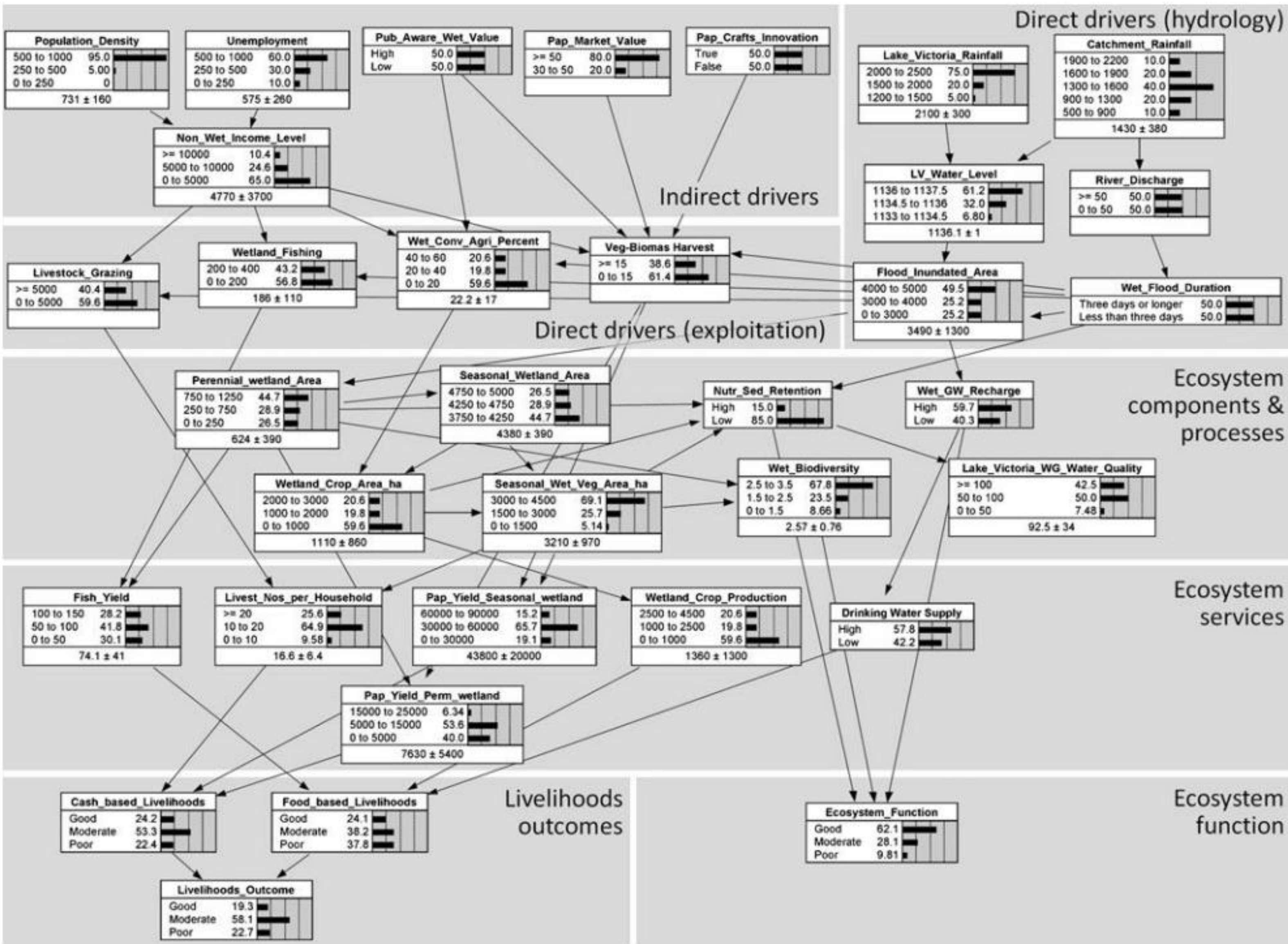
=> Map of habitat suitability for the studied species:



Tantipisanuh et al (2014) “Bayesian networks for habitat suitability modeling: a potential tool for conservation planning with scarce resources” Ecological Applications

Linking Hydrology, Ecosystem Function, and Livelihood Outcomes in African Papyrus Wetlands Using a Bayesian Network Model

Van Dam et al, 2013 Wetlands



Forecasting the population of polar bears

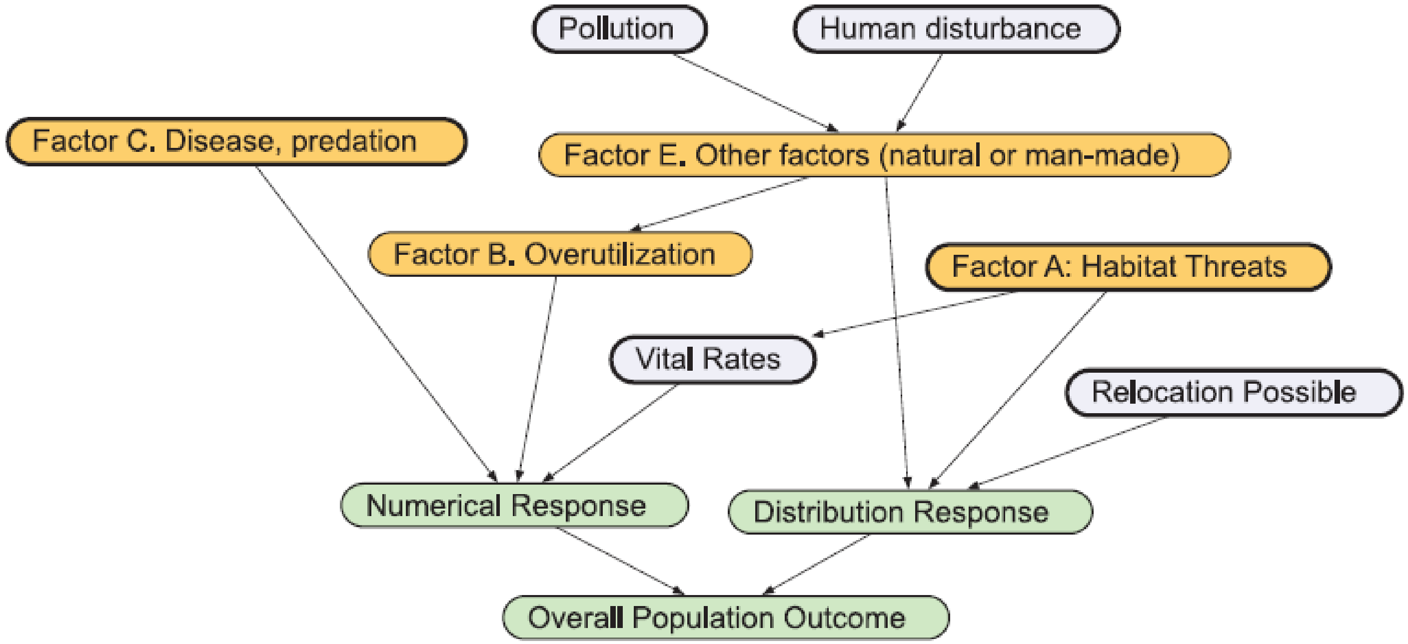
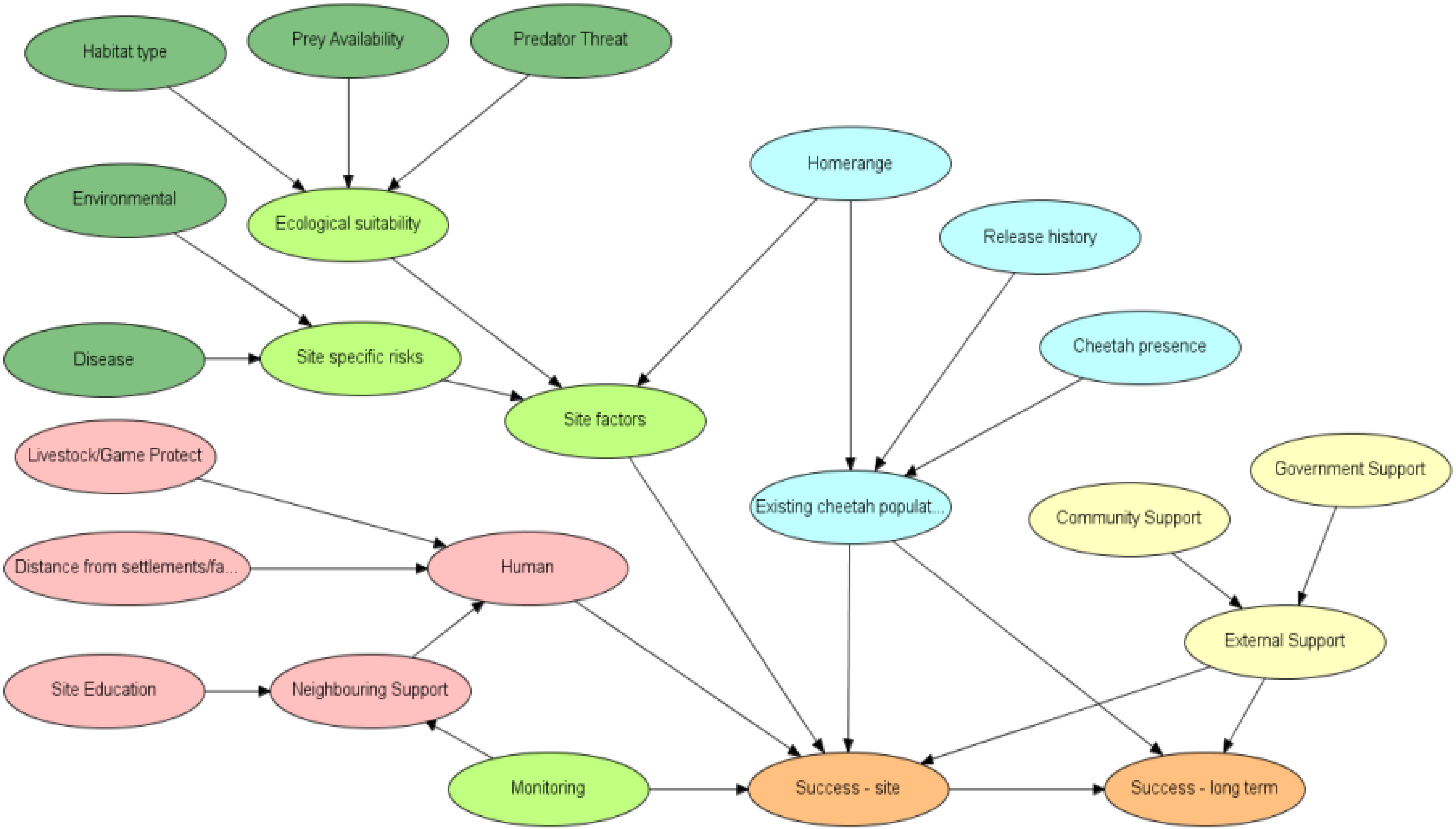


Plate 2. Basic influence diagram for the Bayesian network polar bear population stressor model showing the role of four listing factor categories (orange) used by U.S. Fish and Wildlife Service.

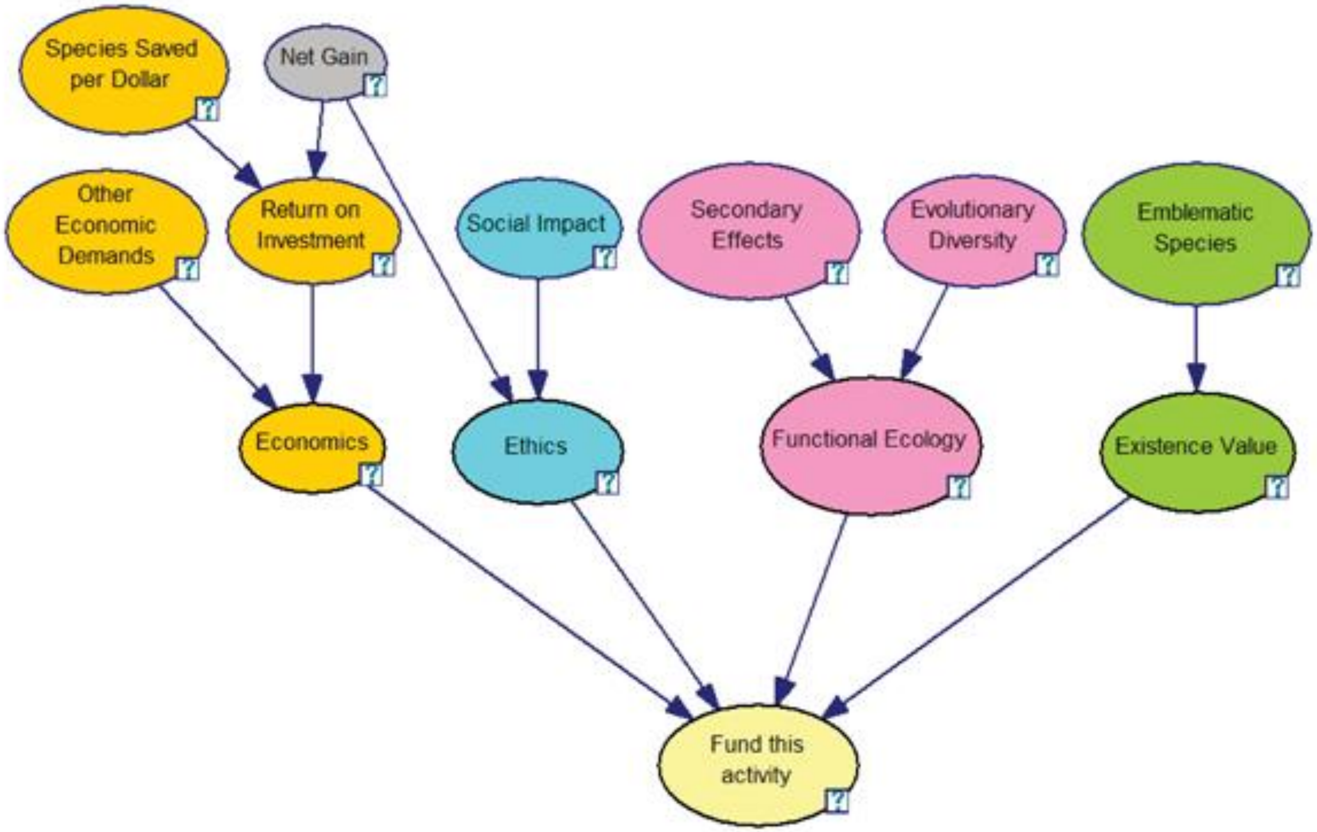
Amstrup, S. C. , B. G. Marcot, and D. C. Douglas. 2008. A Bayesian Network Modeling Approach to Forecasting the 21st Century Worldwide Status of Polar Bears. Pages 213-268 In Eric. T. DeWeaver, Cecilia M. Bitz, and L.-Bruno Tremblay Eds. Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications. Geophysical Monograph 180. American Geophysical Union, Washington DC.

Cheetah relocation success in southern Africa



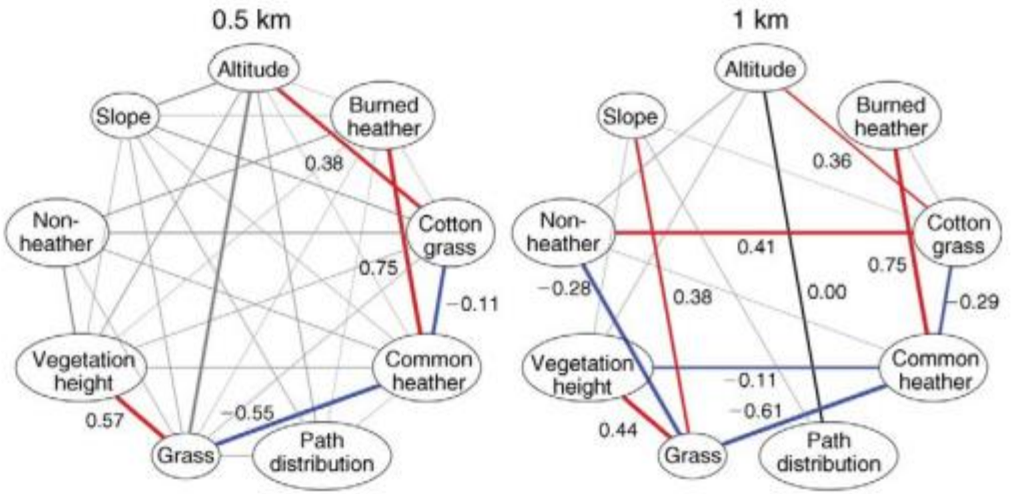
Sandra Johnson. 'Integrated Bayesian Network frameworks for modelling complex ecological issues'. Thesis. 2009

Make decisions about conserving endangered species



Stewart, Gavin, et al. "To Fund or Not to Fund: Using Bayesian Networks to Make Decisions About Conserving Our World's Endangered Species." *CHANCE* 26.3 (2013): 10-17.

Ecological network analysis



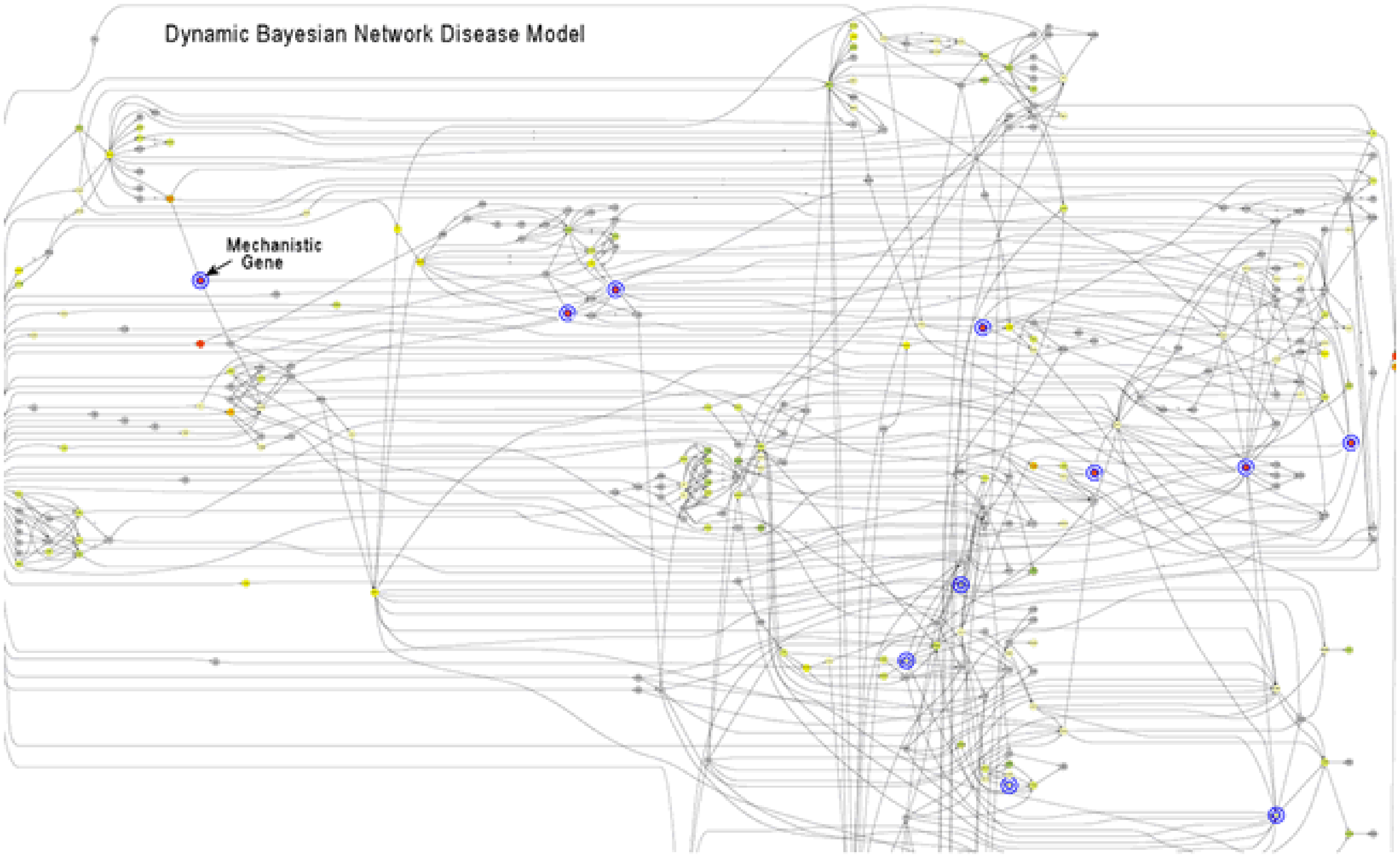
Inter-habitat networks. Networks for 2spatial scales are shown. Habitat variables are enclosed in ovals; lines connecting them represent recovered relationships. Line thickness is scaled to the square of relationship probability. Relationships identified as being highly probable functional relationships are coloured red for positive, blue for negative, and black for non-monotonic, and are labelled with mean influence scores; all other relationships are in grey.

BN shows relationships among habitat types
and interspecific relationships

+ novel insights into ecosystem structure
and identification of key species with high connectivity.

Milns, Isobel, Colin M. Beale, and V. Anne Smith. "Revealing ecological networks using Bayesian network inference algorithms." *Ecology* 91.7 (2010): 1892-1899.

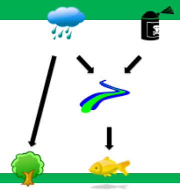
Dynamic BN: linking through time

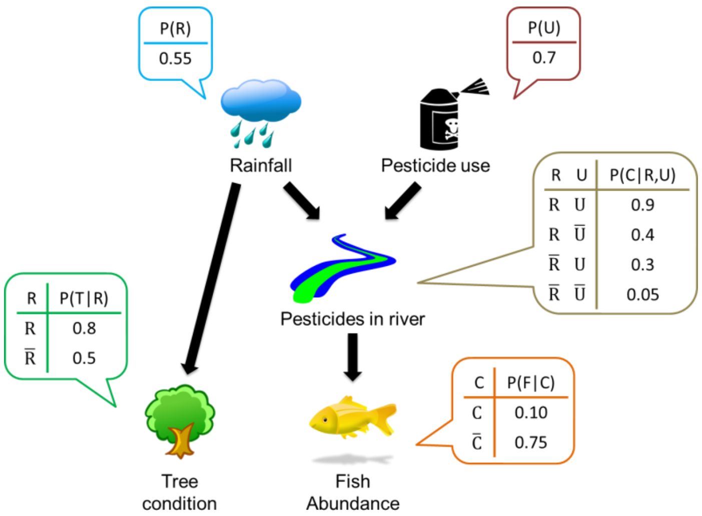


A static Bayesian network does not work for analysing an evolving system that changes over time.

A dynamic BN is used to describe how variables influence each other over time based on the model derived from past data.

Bayesian Networks – Formal definition





1) <u>The graph</u>
Let **G = (V,E)** be a Directed Acyclic Graph (DAG)
V: finite set of nodes (vertices)
E: finite set of edges (arrows) between nodes
The graph defines the structure of the BN

2) <u>Local probabilities</u>
Each node **v** represents a random variable $\mathbf{X_v}$ (or an observation / hypothesis)
Each node **v**, linked to parent(s) **pa(v)**, has a conditional probability table: $\mathbf{P(X_v \mid X_{pa(v)})}$

The set of local probability distributions for all variables in the network is P_X.

→ **A BN for a set of random variables X is the pair (G, P_X)**
A Bayesian network shows conditional probability and possible dependence between the variables.

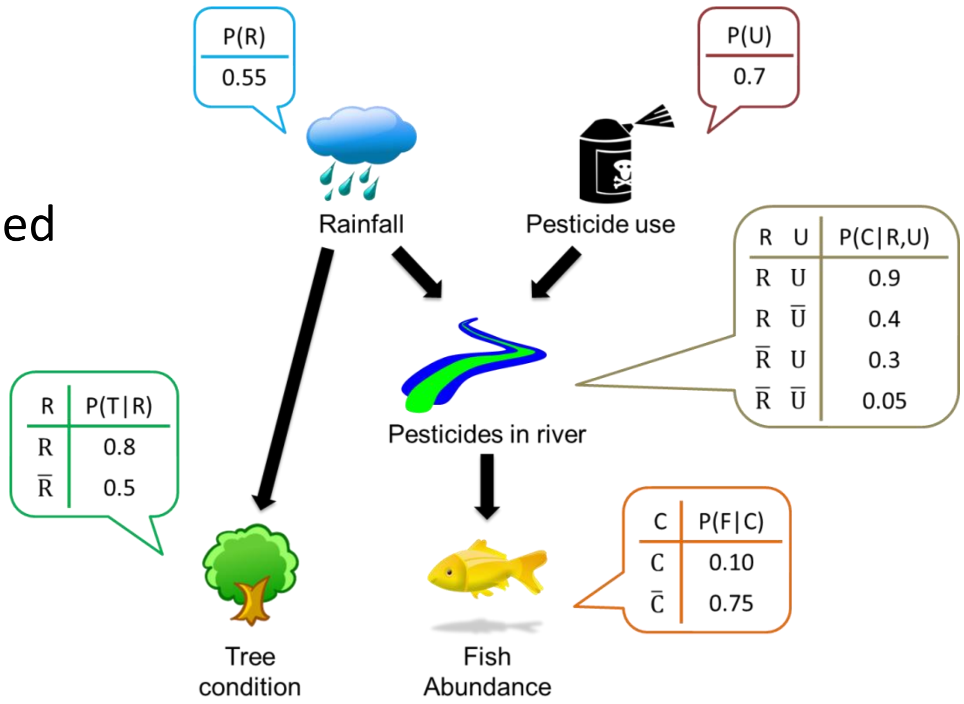
Conditional independence assumptions are implied by the absence of a link.

Bayesian Networks – Formal definition

Parent: node where an arc originates
Child: node where an arc ends

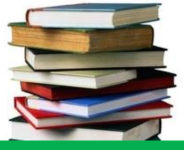
Descendants: nodes that can be reached from a specific node
Ancestors: nodes that lead a path to a specific node

No child can be its own ancestor or descendent → no loop



- The random variables (nodes) can be thought of as states of affairs
  - E.g. True or False
  - Low, medium, high
  - Or continuous (but with discrete states)

Events and Random Variables



- The sample space of a random experiment is a set that includes all possible outcomes
- An event is a subset of outcomes of the experiment.

- A random variable is defined on a set of possible outcomes (the sample space) and a probability distribution that associates each outcome with a probability.
- A random variable represents a measurable aspect or property of the outcomes
- Random variables can be discrete (assume any of a specified list of exact values), or continuous (any numerical value in a interval)

- Example:
- Process of rolling a die.

- Sample space = {1,2,3,4,5,6}

- The random variable X is defined as the number rolled. X=3 is an event

- If X follows a uniform distribution:

$$X(\omega) = \omega$$

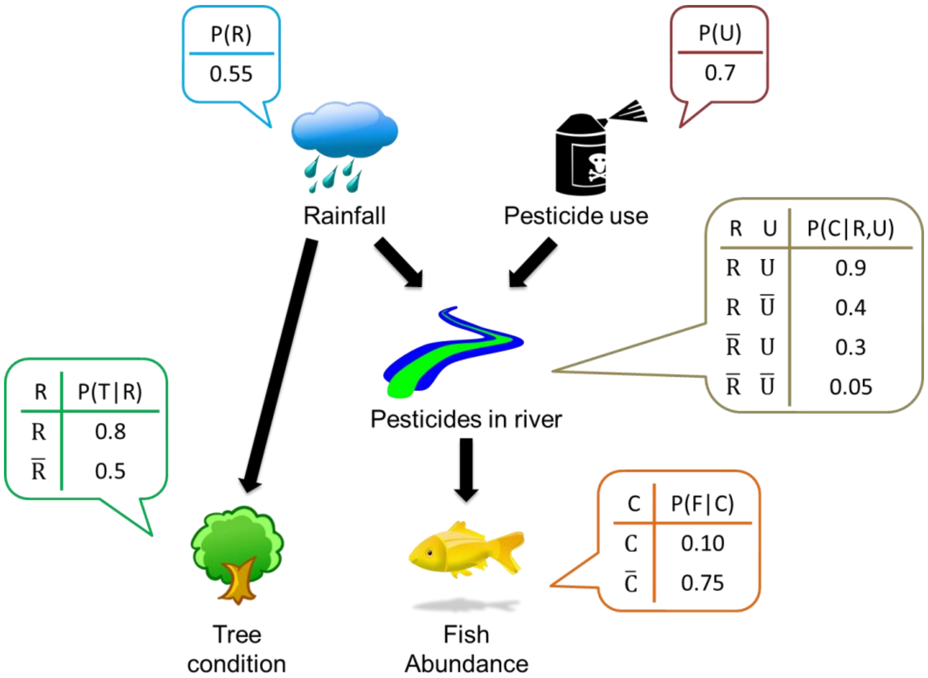
and

$$\rho_X(x) = \begin{cases} \frac{1}{6}, & \text{if } x = 1, 2, 3, 4, 5, 6, \\ 0, & \text{otherwise.} \end{cases}$$

BN – The conditional probability tables (CPT)

- The CPTs contain the P(outcome|cause), i.e. P(child | parents): The probability distribution of a node given its parents.
- The number of columns is the number of states (number of values the random variable can take) (-1) (true/false or high/medium/low, …)
- The number of rows is the product of the number of states for each parents node
  - E.g. R has 2 states, U has 2 states so CPT has 4 rows. If R has 3 states (e.g below avg, avg and above avg) and U has 2 then the CPT of C has 6 rows
  - number of CP (rows) = nb_states(p1) x nb_states(p2) x nb_states(p3).. (all the combinations)
- Rows must total 100% (the sum of probabilities of all possible outcome states for a given set of prior conditions)
- but column totals do not (the sum of likelihoods of all possible prior conditions for a given state).

		P(PesticideInRiver | PesticideUse, Rainfall)	
Pesticide Use	**Annual Rainfall**	**High**	**Low**
High	Below Avg	0.3	0.7
High	Average	0.6	0.4
High	Above Avg	0.8	0.2
Low	Below Avg	0.1	0.9
Low	Average	0.2	0.8
Low	Above Avg	0.3	0.7



BN – The conditional probability tables (CPT)

To specify the probability distribution of a Bayesian network one must give:
- the prior probabilities of all root nodes (nodes with no predecessors)
- the conditional probabilities of all non-root nodes given all possible combinations of their direct predecessors.

The conditional probabilities may be found from observed states or based on:

- a continuous function (then discretized)
P(A|B)=f(B) (function of parent variable)

- a binary logic relationship
P(C|A,B)=¬A OR A AND B
(binary function of parent nodes)

-Or replaced by evidence:
P(A = k)=1

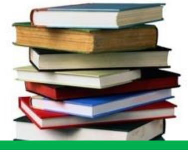
Example: River Flow is given by a Normal distribution with a mean dependent on Drought and Irrigation, and R (rainfall), and a fixed std.

Drought	**Irrigation**	**Mean River Flow**
Yes	Yes	$R/3$
Yes	No	$R/2$
No	Yes	$R/2$
No	No	R

p (RiverFlow | Drought, Rainfall, Irrigation) =
NormalDist(RiverFlow,
Drought==Yes && Irrigation==Yes ? Rainfall/3 :
Drought==Yes && Irrigation==No ? Rainfall/2 :
Drought==No && Irrigation==Yes ? Rainfall/2 :
Rainfall,
50) Discretization is [400; 100; 0] for Good, Poor. These units are arbitrary.

'?:' if ..then..elseif

Independence and conditional independence



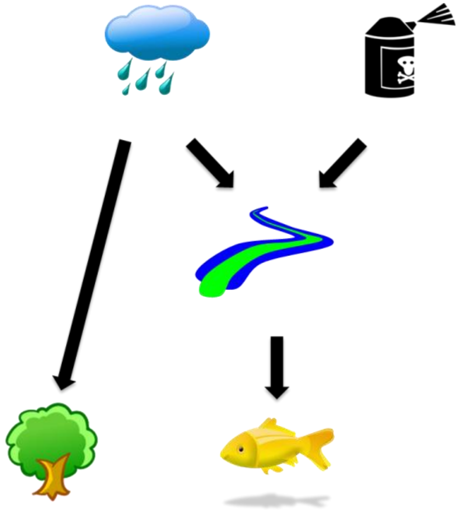
Independence $A \perp\!\!\!\perp B$ $$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B) \Leftrightarrow \mathrm{P}(A \mid B) = \mathrm{P}(A)$$

Mutual independence $$\mathrm{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathrm{P}(A_k)$$

Conditional independence: $A \perp\!\!\!\perp B \mid C$

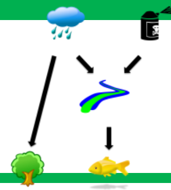
A and B are conditionally independent given C if and only if
$$\mathrm{P}(A \cap B \mid C) = \mathrm{P}(A \mid C) \times \mathrm{P}(B \mid C)$$
$$\Leftrightarrow \mathrm{P}(A \mid B \cap C) = \mathrm{P}(A \mid C)$$

If we know C, knowing B does not bring more information to predict A. E.g. Data of tree condition and fish abundance may have a statistical relationship. They also each have a relationship with rainfall. If we know the rainfall state the tree condition does not help predict the fish abundance. Fish and trees are conditionally independent given rainfall.



-> Facilitates the BN evaluation.

BN – The joint distribution



- “Solving” the BN is calculating the probability of all events concurrently

- The Joint probability is defined as the probability that a series of events happen concurrently.
- The joint distribution of a set of random variables v1, v2, …vn is P(v1,v2,..vn) for all values of v1,v2,…vn

- For binary variables A and B the probabilities of the joint distribution are:
  - $P(A, B)$
  - $P(A, \overline{B})$
  - $P(\overline{A}, B)$
  - $P(\overline{A}, \overline{B})$, and sum to 1

  each can be found with the Chain rule $$P(A \cap B) = P(A \mid B) \times P(B)$$

- General case: The joint distribution of a set of random variables gives the probabilities of all the combinations of the states of the variables (nb states x nb variables)

for n discrete random variables $X_1, X_2, \ldots, X_n$

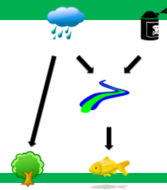
$$\sum_i \sum_j \cdots \sum_k P(X_1 = x_{1i}, X_2 = x_{2j}, \ldots, X_n = x_{nk}) = 1$$

$$\boxed{P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k \,\middle|\, \bigcap_{j=1}^{k-1} A_j\right)}$$

And the general chain rule:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) = {} & P(X_1 = x_1) \\ & \times P(X_2 = x_2 \mid X_1 = x_1) \\ & \times P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \times \cdots \times P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}) \end{aligned}$$

BN – The joint distribution - Compactness

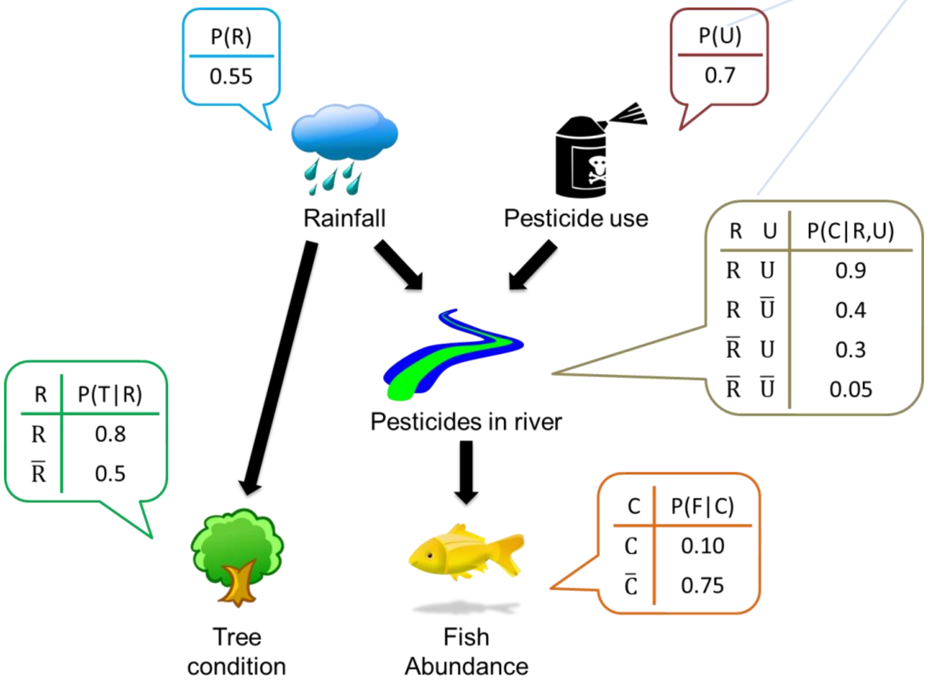


- In our toy example, the joint distribution can be expressed as (chain rule):
$$P(F,T,C,R,U) = P(F|T,C,R,U)P(T|C,R,U)P(C|R,U)P(R|U)P(U)$$

- This simplifies due to conditional independences expressed in the graph:
$$P(F,T,C,R,U) = P(F|C)P(T|R)P(C|R,U)P(R)P(U)$$

We can remove all conditioning nodes except the parents (E.g. if I know the concentration of pesticides in river is high then the state of rainfall does not help to predict the abundance of fish)

If a node does not have a parent, like node R, its probability distribution is unconditional (independent)



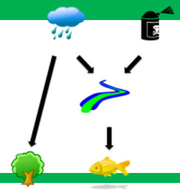
The joint probability of several variables can be calculated from the product of individual conditional probabilities of the nodes.

Only the conditional probabilities specified in the BN are needed (compactness!)

The BN defines a single joint distribution

One joint distribution describes the graph, regardless of the arc direction.

The BN chain rule



→ The set of local probability distributions for all variables in the network P_X factorises over the graph **G**.

$$P(X_1, X_2, \dots, X_n) = \prod_{v=1}^{n} P(X_v | X_{pa(v)})$$

This is called the chain rule for Bayesian networks.

Each factor represents a conditional probability of a node given its parents only (nodes are conditionally independent of their non-descendants given their parents)

(e.g. $\mathrm{P(F,T,C,R,U) = P(F|C)P(T|R)P(C|R,U)P(R)P(U)}$)

→ **A BN for a set of random variables X is the pair (G, P_X)**, where P_X factorises over **G**.

- We can view the graph as encoding a generative sampling process, where the value for each variable (node) is randomly selected using a distribution that depends only on its parents.

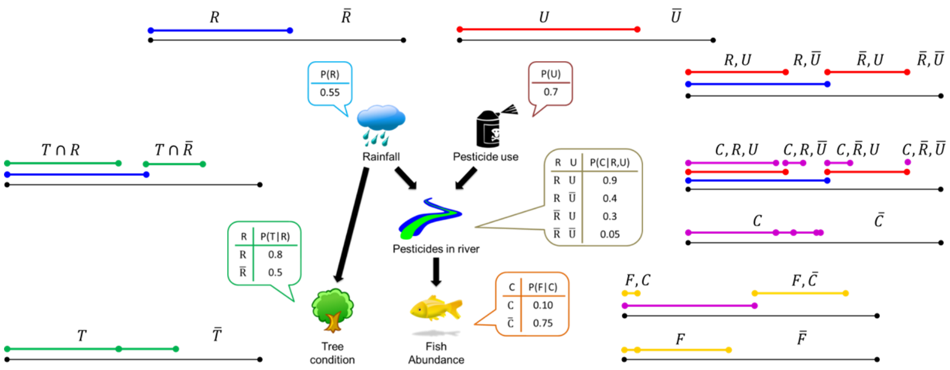
BN: Marginal distribution computation

- What's the probability of high abundance of fish?

-After pruning T, summing the joint distribution where F=1:

$$
\begin{aligned}
P(F) = \;& P(F, C, R, U) && 0.1\text{x}0.9\text{x}0.55\text{x}0.7 \\
+\;& P(F, \bar{C}, R, U) && 0.75\text{x}0.1\text{x}0.55\text{x}0.7 \\
+\;& P(F, \bar{C}, \bar{R}, U) && 0.75\text{x}0.7\text{x}0.45\text{x}0.7 \\
+\;& P(F, \bar{C}, \bar{R}, \bar{U}) \\
+\;& P(F, C, \bar{R}, U) \\
+\;& P(F, C, \bar{R}, \bar{U}) \\
+\;& P(F, C, R, \bar{U}) \\
+\;& P(F, \bar{C}, R, \bar{U}) \\
=\;& 0.39
\end{aligned}
$$

$P(F, \bar{C}, \bar{R}, U)$
$=$
$P(F|\bar{C})P(\bar{C}|\bar{R}, U)P(\bar{R})P(U)$



Each probability can be computed from the BN chain rule:

$$P(F, C, R, U) = P(F|C)P(C|R, U)P(R)P(U)$$

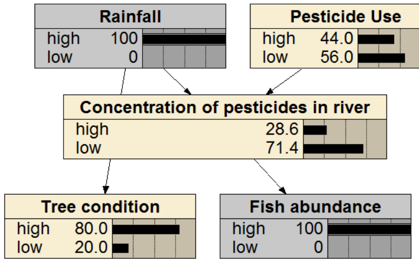
-Or recursively:

$P(F) = P(F|C)P(C) + P(F|\bar{C})P(\bar{C})$

$P(C) = P(C|R, U)P(R, U) + P(C|\bar{R}, U)P(\bar{R}, U) + P(C|R, \bar{U})P(R, \bar{U}) + P(C|\bar{R}, \bar{U})P(\bar{R}, \bar{U})$

$P(R, U) = P(R)P(U)$

BN - Queries

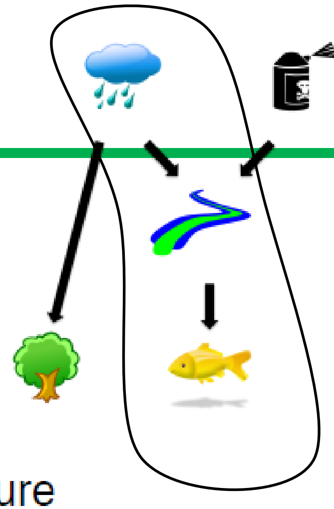


- Queries to BNs
  - Probability conditioned on evidence: **$P(X_v \mid E = e)$**
    - E.g. P(C | F=high) or P(R,U,C,T | F=low)
    - **$P(X_v \mid E = e) = P(X_v, E = e)/P(E = e)$**
  - The most probable explanation (MPE)
    - most likely state of all non observed variables
  - The maximum a posteriori (MPA)
    - Most likely state of a subset of non observed variables given the evidence

- Simple method to answer these queries:
  - generate the joint distribution
  - and exhaustively sum out the joint distribution (in the case of a conditional probability query)
  - search for the most likely entry (in the case of an MPE query)
  - or both (in the case of an MAP query)

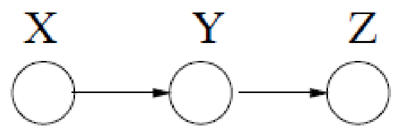
  - However this is inefficient for large BNs where approximate inference and optimisation should be used.
  - Several methods / software to compute optimum approximations, instead of exact solution.

BN – Flow of information



Built-in independence assumptions

A simple case: Indirect connection



- Think of X as the past, Y as the present and Z as the future
- This is a simple **Markov chain**
- We interpret the lack of an edge between X and Z as a conditional independence, $X \perp\!\!\!\perp Z | Y$. Is this justified?
  - Based on the graph structure, we have: BN chain rule:

$$p(X, Y, Z) = p(X)p(Y|X)p(Z|Y)$$

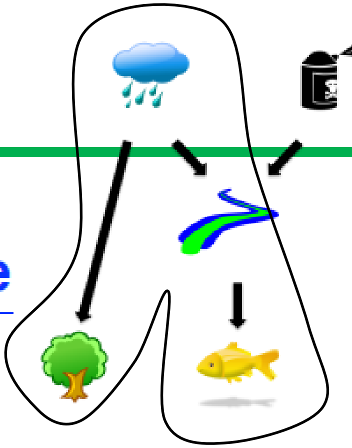
  - Hence, we have: General chain rule:

$$p(Z|X, Y) = \frac{p(X, Y, Z)}{p(X, Y)} = \frac{p(X)p(Y|X)p(Z|Y)}{p(X)p(Y|X)} = p(Z|Y)$$

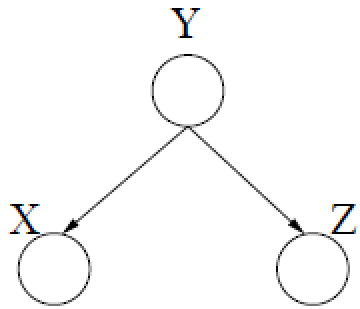
  - Note that the edges that are *present* do *not* imply dependence. But the edges that are *missing* do imply independence.

Machine Learning course by Doina Precup: http://www.cs.mcgill.ca/~dprecup/courses/ML/lectures.html

BN – Flow of information



A more interesting case: Common cause



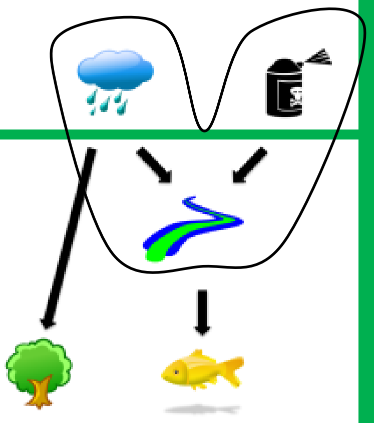
- Again, we interpret the lack of edge between X and Z as $X \perp\!\!\!\perp Z | Y$. Why is this true?

$$p(X|Y,Z) = \frac{p(X,Y,Z)}{p(Y,Z)} = \frac{p(Y)p(X|Y)p(Z|Y)}{p(Y)p(Z|Y)} = p(X|Y)$$

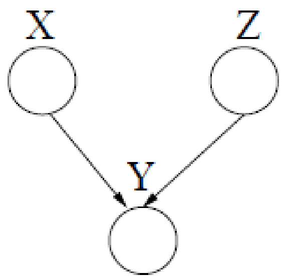
- This is a “hidden variable” scenario: if Y is unknown, then X and Z could appear to be dependent on each other

Machine Learning course by Doina Precup: http://www.cs.mcgill.ca/~dprecup/courses/ML/lectures.html

BN – Flow of information



The most interesting case: V-structure



- In this case, the lacking edge between X and Z is a statement of *marginal independence*: $X \perp\!\!\!\perp Z$.
- In this case, once we know the value of Y, X and Z might depend on each other.
- E.g., suppose X and Z are independent coin flips, and Y is true if and only if both X and Z come up heads.
- Note that in this case, X is *not* independent of Z given Y!
- This is the case of "explaining away".

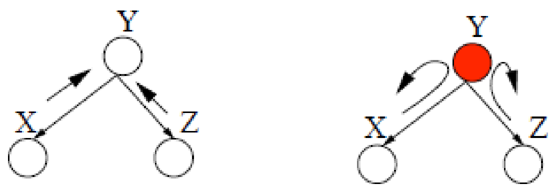
BN – Flow of information

Evidence blocks the flow of influence in a linear paths and in common cause path, but binds the causes in a common effect path

- *Head-to-tail*

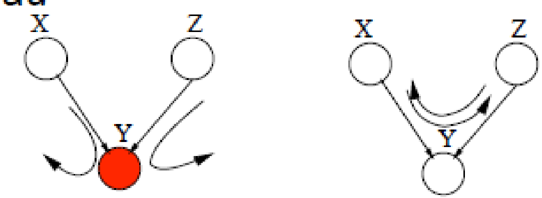
Y unknown, path unblocked Y known, path blocked

- *Tail-to-tail*



Y unknown, path unblocked Y known, path blocked

- *Head-to-head*

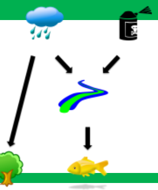


Y unknown, path BLOCKED Y known, path UNBLOCKED

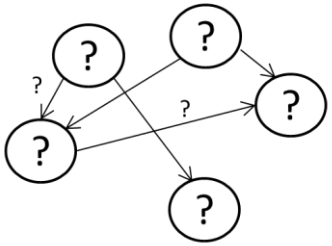
As evidence comes in, it is tempting to think of the probabilities of the nodes changing, but, what is changing is the conditional probability of the nodes given the changing evidence.

Machine Learning course by Doina Precup: http://www.cs.mcgill.ca/~dprecup/courses/ML/lectures.html

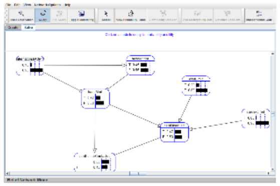
BN next!

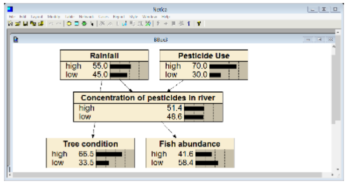


- we know how to solve small BNs by hand



- We will see:
  - How to develop BNs
  - The tools available to solve complex BNs





- Then we will detail a BN used in a current research work in Silwood: Terragenesis