Reasoning and Decision-Making under Uncertainty

6. Session:
Modeling & Reasoning with Bayesian Networks

Prof. Dr.-Ing. Stefan Kopp
Center of Excellence „Cognitive Interaction Technology“
AG Sociable Agents

CITEC

Sociable Agents

Using Bayesian networks

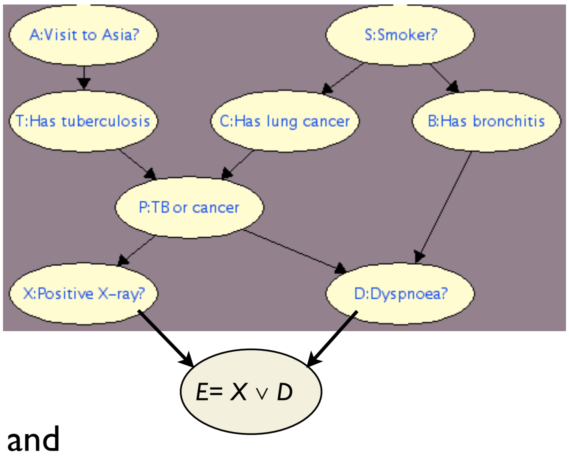
Applying Bayesian networks to real-world problems requires two steps:

1. construct an „appropriate“ Bayesian network model for a domain
   - variables ➔ capture world in terms of states and events
   - graph structure ➔ capture dependencies/causal structure of the world
   - parameters (CPTs) ➔ capture correlations and contingencies

2. draw required inferences by applying appropriate queries
   - prob. of evidence: $Pr(\mathbf{e})=?$
   - prior/post. marginals: $Pr(x_1,...,x_m|\mathbf{e})=?$
   - most probable explanation (MPE): $\mathbf{x}$=? with $Pr(x_1,...,x_n|\mathbf{e})=max$
   - maximum aposterior hypothesis (MAP): $\mathbf{x}$=? *with* $Pr(x_1,...,x_m|\mathbf{e})=max$

Probability of evidence

Query: How likely is some variable instantiation **e** ➔ *Pr(***e***)=?*



Example: *Pr(X=yes, D=no)=?*

Example: *Pr(X=yes ∨ D=yes)=?*

can be computed indirectly with the auxiliary-node technique:

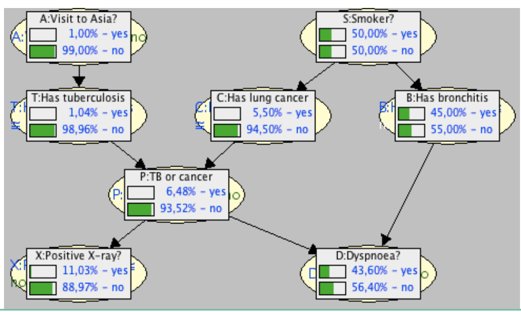
- add node *E* with *X,D* as parents and *Pr(e|x,d)=1 iff e=1 and (d=1 or x=1)*
- possible when not too many evidence var's

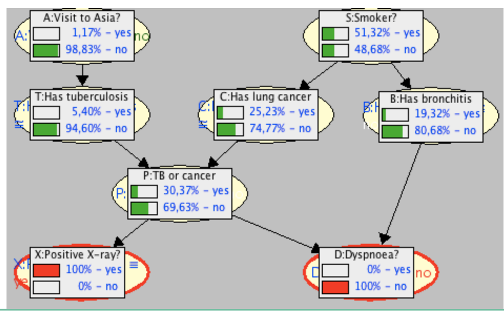
Prior and posterior marginals

Query: What is the probability distribution for a limited set of variables with (posterior) or without (prior) some given evidence?
➔ *Pr(x1,...,xm)=?* or *Pr(x1,...,xm|***e***)=?*

Definition: Given joint distribution $Pr(x_1,...x_n)$ and number *m* of variables

- prior marginal : $$Pr(x_1,...,x_m) = \sum_{x_{m+1},...x_n} Pr(x_1,...,x_n)$$
- posterior marginal given **e** : $$Pr(x_1,...,x_m|\mathbf{e}) = \sum_{x_{m+1},...x_n} Pr(x_1,...,x_n|\mathbf{e})$$

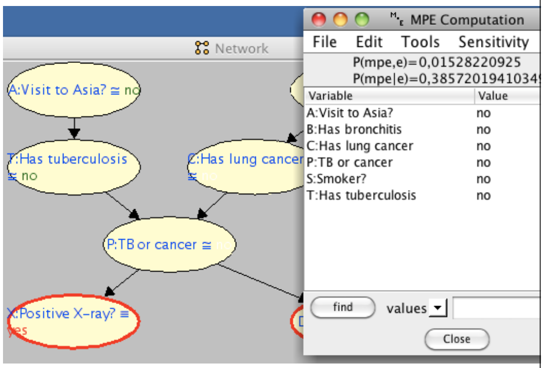




Most probable explanation (MPE)

Query: What is the most probable instantiation of *all* network var's, given some evidence **e** → ***x*** *with* $Pr(x_1,...,x_n|\mathbf{e})=max$?

Example: MPE for positive x-ray and not dyspnoea?



Cannot be computed directly from the maximal posterior marginals

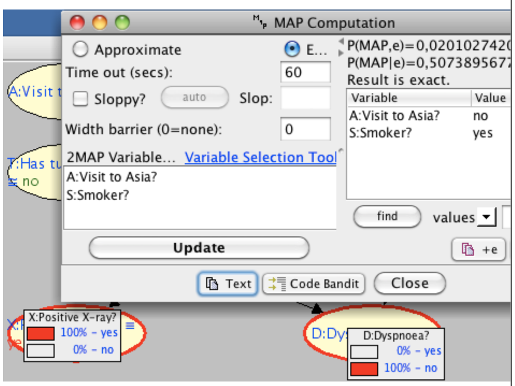
- choosing x_i such that $Pr(x_i|\mathbf{e})=max$ yields expl. *p* with *smoker=true and* $Pr(p|e)=20.03\%$ whereas $Pr(mpe|e)=38.57\%$

Maximum a posteriori hypothesis (MAP)

Query: What is the most probable instantiation of a subset of var's $\mathbf{M}=X_1,...,X_m$ given some evidence **e** → **m** *with* $Pr(\mathbf{m}|\mathbf{e})=max$?

- MPE is a special case of MAP, easier to compute algorithmically

Example: Given *X=yes, D=no,* what is the most probable instantiation of **M**={A,S}?



Approximative method:

- compute MPE and return values for MAP variables (projecting MPE on MAP var's)
- but, leads to *A=no, S=yes* here with prob ~48%, while *A=no, S=no* is MAP with prob ~50%

Bayesian models

Diagnostic models vs. causal models

- so far, we have introduced Bayesian networks as causal models
- directed links with parents = causes, children = direct effects

Alternative option: diagnostic model

- links from symptoms to explanations, i.e. with prob's for explanations conditioned upon symptoms
- requires additional and strange dependencies between otherwise independent causes and often between separately occurring symptoms

Causal models are usually preferable as they require fewer parameters, and numbers that are easier to come up with.

Bayesian models

How to construct a Bayesian network?

1. define network variables and their values
   - distinguish between *query*, *evidence*, and *intermediary* variables
   - query and evidence var's usually determined from problem statement
   - intermediary (a.k.a. *hidden* or *latent)* variables often less obvious

3. define network structure
   - for each var *X*, answer the question: what set of var's are direct causes of *X*?

5. define network parameters (CPTs)
   - difficulty and objectivity depend on problem and available data
   - often assuming a distribution (model) and estimate parameters

Example I:

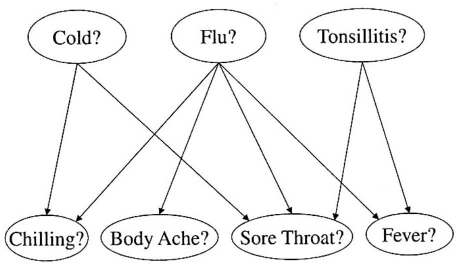
„Flu is an acute disease characterized by fever, body aches, and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a soar throat. Tonsillitis is an inflammation of the tonsils that leads to a soar throat and can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

Variables:

- *query: flu, cold, tonsillitis*
- *evidence: chilling, body ache and pain, sore throat, fever*
- *intermediary*: -
- values: {true,false}

Structure?

Example I:

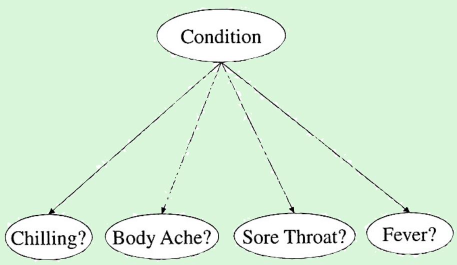


CPTs normally obtained from experts (subjective beliefs, empirical data)

- problem of parameter estimation
- *Example:* Given N patient records $\mathbf{d}_i$, find parametrization Θ such that

$$\prod_{i=1}^{N} Pr(\mathbf{d}_i) = max$$

Naive Bayes structure



- class variable *Condition* ∈ {*normal, cold, flu, tonsillitis*}
- attributes *Chilling, Body Ache, ...*
- single-fault assumption: only one cond. can hold at any time
- inconsistent with info: given *Cond.=Cold, Fever* and *Sore Throat* would become independent

Example II:

„Few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is scanning with a false positive of 1% and a false negative of 10%. The second is a blood test of progesterone with a false positive of 10% and a false negative of 30%. The third is a urine test of progesterone with false positive of 10% and a false negative of 20%. The prob. of a detectable progesterone level is 90% given pregnancy and 1% given no pregnany. The prob. that insemination will impregnate a cow is 87%.“

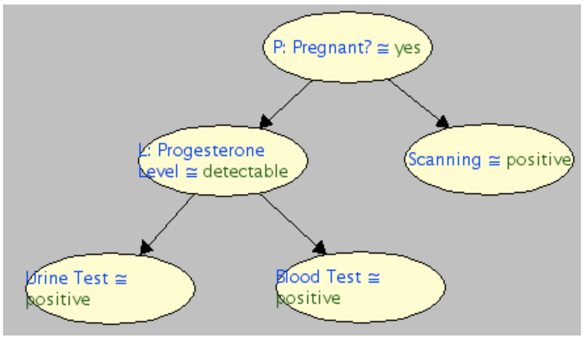
Goal: Build network to compute prob of pregnany given some test results

Variables:

- *query*: pregnancy? (*P*)
- *evidence*: scanning (*S*), blood test (*B*), urine test (*U*)
- *intermediary*: progesterone level (*L*)

Example II:

Structure:

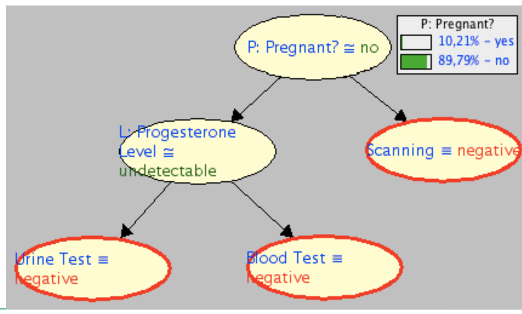


CPTs directly given by problem statement, e.g.

P	*L*	*P(l|p)*
yes	*undetect.*	*0.1*
no	*detectable*	*0.01*

Example: After insemination, all three tests are negative.

- *Pr(P|***e***)=?*
  Still 10,21%



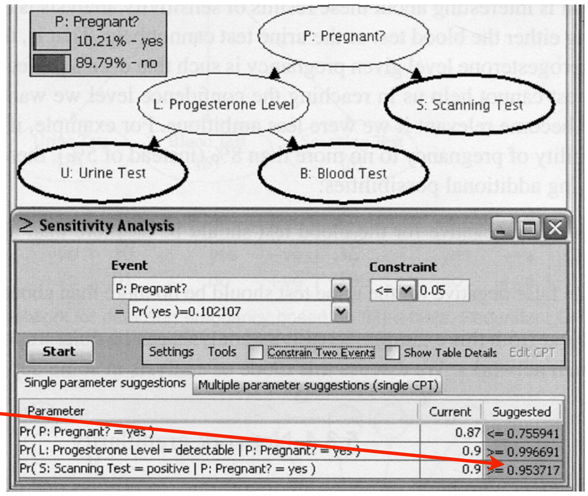
Example II: sensitivity analysis

Q: What kind of a test is needed to get this error prob. down to ~5%?

- acceptable false positive/false negative rates?

Sensitivity analysis:
which network parameters do we have to change, and how much, in order to ensure that *Pr(P|L=neg.,B=neg.,U=neg.)≤5%* ?

- what helps is to improve the scanning test to a false negative of 4,63%



Inference in Bayesian Networks

How to implement these inferences? Need technique to calculate/update all probabilities in the network given some evidence (also required for MPE and MAP).

- Exact algorithms
  - variable elimination and factor elimination (marginalization, enumeration)
  - jointree algorithm
  - (recursive) conditioning

- Approximative algorithms
  - belief propagation
  - stochastic sampling
  - Monte Carlo Markov Chain

Inference by variable elimination

How to compute prior marginals and probability of evidence?

Variable elimination:

- given a distribution *Pr(A,B,C,D,E)*, variable *A* with values a_i can be „summed out" („marginalized") by
- *Example:*

$$Pr(B,C,D,E) = \sum_{a_i} Pr(a_i, B, C, D, E\ )$$

A	*B*	*C*	*D*	*E*	*Pr(.)*
true	true	true	true	true	0.063
false	true	true	true	true	0.19

⇨

B	*C*	*D*	*E*	*Pr(.)*
true	true	true	true	0.063+0.19=0.083

- reduces *Pr* from 32(=2^5) to *Pr'* with 16(=2^4) rows
- *Pr'* as good as *Pr* for all queries not related to *A*, but more efficient.

Inference by variable elimination

Sometimes var's can be summed out without having to construct the full distribution, but keeping it in a „factored" form given by Bayes Network

- calculate/update local CPTs
- allows to escape exponential complexity

Definition: factor f over var's **X** is a function that maps each instantiation **x** of **X** to a number f(**x**)≥0

- can represent marginal or conditional distributions over **X**
- called trivial when defined over empty set of variables T

There are two key operations on factors:

- summing out variables (~marginalizing)
- multiplying two factors (~chain rule)

Inference by variable elimination

Summing out variables

Definition: factor f over **X** and $X \in$ **X**. The result of summing out var X from f is another factor over var's **Y**=**X**\{X} defined as $(\sum_X f)(\mathbf{y}) := \sum_x f(x, \mathbf{y})$

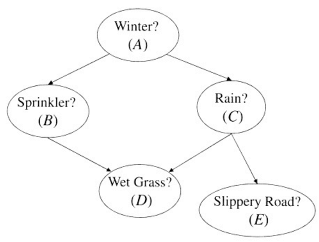
- commutative: $\sum_X \sum_Y f = \sum_Y \sum_X f$
- also called marginalizing variables **X** or projecting *f* on variables **Y**
- size *O(exp(w))* for *w* = #var's in resulting factor

Multiplying factors

Definition: the result of multiplying factors f1(**X**) and f2(**Y**) is a another factor over var's **Z**=**X** ∪ **Y** defined by $(f_1 f_2)(\mathbf{z}) := f_1(\mathbf{x}) f_2(\mathbf{y})$ with $\mathbf{x} \sim \mathbf{z}, \mathbf{y} \sim \mathbf{z}$

- commutative and associative
- size *O(m exp(w))* for *w* = #var's in resulting factor, *m* = #factors multiplied

Inference by variable elimination



Using elimination for inference

- view CPTs as factors and express joint distribution through factor multiplication (chain rule)

$$Pr(a,b,c,d,e) = \Theta_{E|C}\Theta_{D|BC}\Theta_{C|A}\Theta_{B|A}\Theta_A$$

- compute marginal distribution by summing out variables from this product

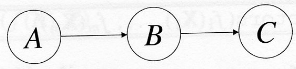
$$Pr(D,E) = \sum_{A,B,C} \Theta_{E|C}\Theta_{D|BC}\Theta_{C|A}\Theta_{B|A}\Theta_A$$

- answer queries from these marginal distributions

Easy, but complex. Can be optimized as $\sum_X f_1 f_2 = f_1 \sum_X f_2$ if X appears only in f_2

- to sum out *X* from product, need to multiply only those factors that included *X* before summing out!

Example: compute prior marginal *Pr(C)* by eliminating first *A*, then *B* from chain product



- $Pr(C) = \underbrace{\sum_B \Theta_{C|B} \underbrace{\sum_A \Theta_A \Theta_{B|A}}}$

A	Θ_A
true	.6
false	.4

A	B	$\Theta_{B\|A}$
true	true	.9
true	false	.1
false	true	.2
false	false	.8

B	C	$\Theta_{C\|B}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

A	B	$\Theta_A\Theta_{B\|A}$
true	true	.54
true	false	.06
false	true	.08
false	false	.32

B	$\sum_A \Theta_A\Theta_{B\|A}$
true	$.62 = .54 + .08$
false	$.38 = .06 + .32$

B	C	$\Theta_{C\|B}\sum_A \Theta_A\Theta_{B\|A}$
true	true	.186
true	false	.434
false	true	.190
false	false	.190

C	$\sum_B \Theta_{C\|B}\sum_A \Theta_A\Theta_{B\|A}$
true	.376
false	.624

Order of elimination irrelevant for result, but not for computational costs!

Other possibility: first *B* then *A*

- $Pr(C) = \sum_A \Theta_A \underbrace{\sum_B \Theta_{B|A}\Theta_{C|B}}$

largest factor with 2^3 param's

Best order: smallest possible width = num. of var's in the largest factor constructed

- can be determined, but NP-hard
- heuristics used to generate relatively good orders (see Darwiche, sect. 6.6)

Inference by variable elimination

How to compute posterior marginals?

- need to compute the factor *Pr*(**Q**|**e**) = *Pr*(**Q**,**e**) / *Pr*(**e**)

Compute joint marginal *Pr*(**Q**, **e**) and normalize to get *Pr*(**Q**|**e**)

- gives also *Pr*(**e**) for free since $Pr(\mathbf{e}) = \sum_q Pr(\mathbf{q}, \mathbf{e})$

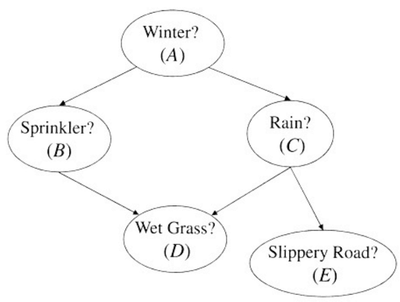
Example: **Q**={*D*,*E*}

D	*E*	*Pr*(**Q**|**e**)
true	true	.448
true	false	.192
false	true	.112
false	false	.248

D	*E*	*Pr*(**Q**,**e**)
true	true	.21504
true	false	.09216
false	true	.05376
false	false	.11904

=.11904 / .48

∑=.48 *Pr*(**e**)



Inference by variable elimination

Zero out all rows incompatible with **e** and use elimination for computing joint marginals (➔ reasoning more efficient with evidence)

Definition: the reduction of factor *f(**X**)* given evidence **e** is another factor over **X** denoted by f^e, defined by

$$f^{e}(\mathbf{x}) := \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \sim \mathbf{e} \\ 0 & \text{otherwise} \end{cases}$$

- it holds: $(f_1 f_2)^{\mathbf{e}} = f_1^{\mathbf{e}} f_2^{\mathbf{e}}$

The joint marginal *Pr(**Q**,**e**)* hence can be computed as follows:

- *Example*: previous network, **Q**={*D*,*E*}

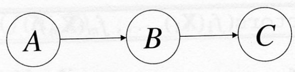
$$Pr(\mathbf{Q},\mathbf{e}) = \sum_{A,B,C} \Theta^e_{E|C}\Theta^e_{D|BC}\Theta^e_{C|A}\Theta^e_{B|A}\Theta^e_{A}$$

Example: compute posterior marginal *Pr(**Q**={C},**e**:A=true)* by eliminating first *A*, then *B*

- $Pr(\mathbf{Q},\mathbf{e}) = \sum_B \sum_A \Theta^e_A \Theta^e_{B|A}\Theta^e_{C|B}$
  $= \sum_B \Theta^e_{C|B}\Theta^e_A\Theta^e_{B|A}$

Therefore:

- *Pr(C=true,A=true)=.192*
- *Pr(C=false,A=true)=.408*
- *Pr(A=true)=.6*
- *Pr(C=true|A=true)=.192 / .6=.32*



A	Θ_A
true	.6
false	.4

A	B	$\Theta_{B\mid A}$
true	true	.9
true	false	.1
false	true	.2
false	false	.8

B	C	$\Theta_{C\mid B}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

A	Θ^e_A
true	.6

A	B	$\Theta^e_{B\mid A}$
true	true	.9
true	false	.1

B	C	$\Theta^e_{C\mid B}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

A	B	$\Theta^e_A\Theta^e_{B\mid A}$
true	true	.54
true	false	.06

B	$\sum_A \Theta^e_A\Theta^e_{B\mid A}$
true	.54
false	.06

B	C	$\Theta^e_{C\mid B}\sum_A \Theta^e_A\Theta^e_{B\mid A}$
true	true	.162
true	false	.378
false	true	.030
false	false	.030

C	$\sum_B \Theta^e_{C\mid B}\sum_A \Theta^e_A\Theta^e_{B\mid A}$
true	.192
false	.408

Inference by factor elimination

One can generalize variable elimination to factor elimination, i.e. elimination of *sets of variables* (Lauritzen & Spiegelhalter 1988).

Basic idea: Want to compute the prior marginal over variable Q

- variable elimination: eliminate other var's from the network
- factor elimination: eliminate factors (with several variables) except one that contains Q, used to answer query

input:
$\mathcal{N}$: a Bayesian network
Q: a variable in network $\mathcal{N}$
output: the prior marginal $\Pr(Q)$
main:
$\mathcal{S} \leftarrow$ CPTs of network $\mathcal{N}$
$f_r \leftarrow$ a factor in $\mathcal{S}$ that contains variable Q
while $\mathcal{S}$ has more than one factor **do**
  remove a factor $f_i \neq f_r$ from set $\mathcal{S}$
  $\mathbf{V} \leftarrow$ variables that appear in factor f_i but not in $\mathcal{S}$
  $f_j \leftarrow f_j \sum_{\mathbf{V}} f_i$ for some factor f_j in $\mathcal{S}$
end while
return project(f_r, Q)

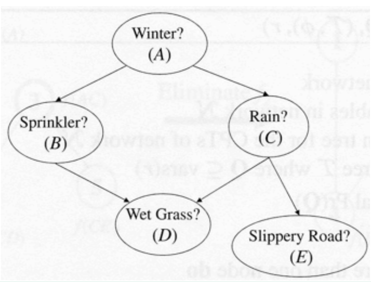
Choice 1: which factor to eliminate?
Choice 2: which variable to eliminate?
Any set of choices valid, but some are more efficient!

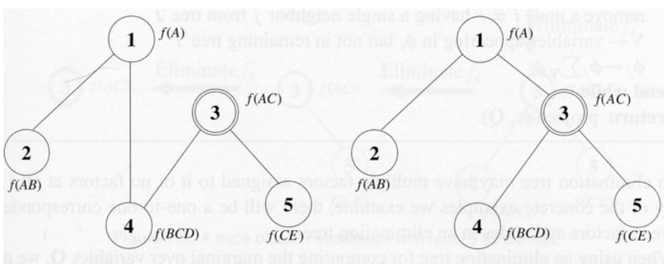
Inference by factor elimination

Elimination order becomes „elimination tree" = organization of factors

Definition: An elimination tree (T, Θ) for a set of factors ***S*** is a *tree T*, in which each factor in ***S*** is assigned to exactly one node in *T*, where Θ_i is the *product* of all factors assigned to node *i* in tree *T*

- factors are CPTs in the Bayesian network
- nodes may have multiple factors assigned to them (or no factors at all)
- different tree structures and correspondences to the network possible





Inference by factor elimination

<u>*Using factor elimination for computing marginal over **Q***</u>

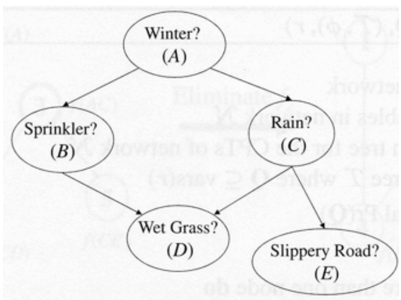
Pick one node *r* with **Q** ⊆ *vars(r)* as root node

<u>*Elimination strategy:*</u>

- pick factor Θ_i only if *i*≠*r* and has a single neighbor *j*
- sum out variables ***V*** that appear in Θ_i but not in rest of the tree
- multiply the result $\sum_V \Theta_i$ into factor Θ_j associated with neighbor *j*
- after eliminating all nodes *i*≠*r*, project factor Θ_r on variables **Q** yields answer to query

<u>Note</u>: Any elimination order (≙elimination tree) will lead to correct results, yet some lead to less work than others

Factor elimination



<u>*Example*</u>: Compute prior marginal over *C*

- four elimination steps: