# COSC1125/1127 Artificial Intelligence

## Week 10: Bayesian Networks

[RN2] Section 14.1-14.4
[RN3] Section 14.1-14.4

# House keeping

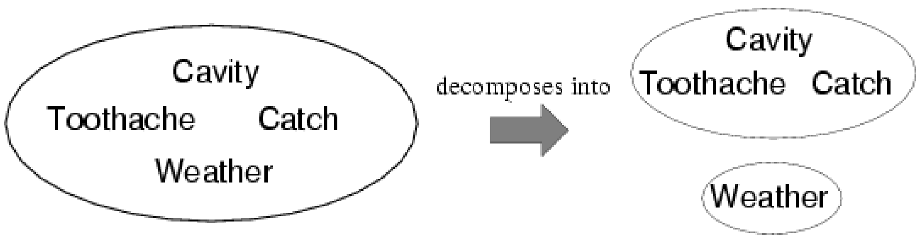
- Assignment#2 marking is almost done.
- Class test results have been released. Please learn from this experience, identifying gaps in understanding and seeking further improvement.
- Getting started on your Assignment#3 (submission instructions have been revised, to request submissions of 3 separate files, NOT a single .zip file).
- We are working on revising tutorial sample solutions, to give you better guidance in terms of what areas you need to focus on. Some tutorial questions will be clearly marked as “optional”.
- CES (Course Experience Survey) is now open: https://surveys.rmit.edu.au/Blue/

  Please provide your feedback for us to improve the running of this course.
- We understand some of you are anxious with the amount of contents covered in this course. A simple rule is that you should follow lecture slides and tutorial exercises. If a topic is not covered in lectures, they will not be examinable in the final exam.
- We will also provide revision tips and go over a past exam, to allow you spend your time wisely for preparation of the final exam.
- There will be **focus sessions** scheduled, to provide additional assistance during the revision period (week 13 – 14).

# Independence

$A$ and $B$ are independent iff

$$\mathbf{P}(A \mid B) = \mathbf{P}(A) \quad \text{or } \mathbf{P}(B \mid A) = \mathbf{P}(B) \quad \text{or } \mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$$




$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$

32 entries reduced to 12; for $n$ independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare.

Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

# Conditional Independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$

The same independence holds if I haven't got a cavity:

(2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$

*Catch* is **conditionally independent** of *Toothache* given *Cavity*:

$\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$

Equivalent statements:

$\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$

$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)$

# Conditional Independence...

Write out full joint distribution using chain rule:

$$\mathbf{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity})$$
$$= \mathbf{P}(\mathit{Toothache} \mid \mathit{Catch}, \mathit{Cavity})\, \mathbf{P}(\mathit{Catch}, \mathit{Cavity})$$
$$= \mathbf{P}(\mathit{Toothache} \mid \mathit{Catch}, \mathit{Cavity})\, \mathbf{P}(\mathit{Catch} \mid \mathit{Cavity})\, \mathbf{P}(\mathit{Cavity})$$
$$= \mathbf{P}(\mathit{Toothache} \mid \mathit{Cavity})\, \mathbf{P}(\mathit{Catch} \mid \mathit{Cavity})\, \mathbf{P}(\mathrm{Cavity})$$

i.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in $n$ to linear in $n$.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

# Bayes' Rule

Product rule $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

$\Rightarrow$ **Bayes' rule**: $$P(a \mid b) = \frac{P(b \mid a)\ P(a)}{P(b)}$$

or in distribution form

$$\mathrm{P}(Y \mid X) = \frac{\mathrm{P}(X \mid Y)\ \mathrm{P}(Y)}{\mathrm{P}(X)} = \alpha\, \mathrm{P}(X \mid Y)\mathrm{P}(Y)$$

Useful for assessing *diagnostic* probability from *causal* probability:

$$\mathrm{P}(Cause \mid Effect) = \frac{\mathrm{P}(Effect \mid Cause)\ \mathrm{P}(Cause)}{\mathrm{P}(Effect)}$$

E.g., let *M* be meningitis, *S* be stiff neck:

$P(m|s) = P(s|m)\, P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$

Note: posterior probability of meningitis still very small!

# Bayes' Rule and Conditional Independence

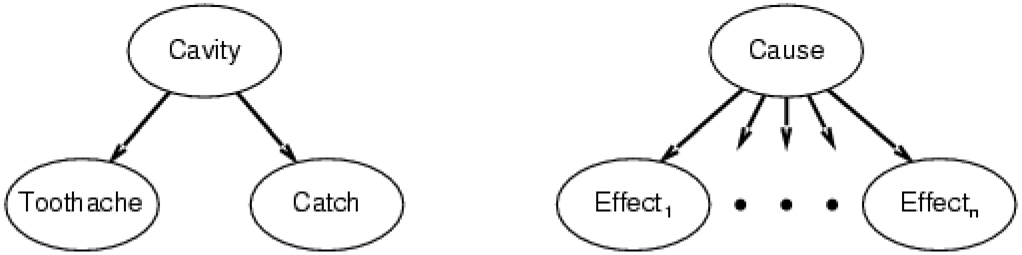
$\mathbf{P}(Cavity \mid toothache \wedge catch)$

$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

$= \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$

This is an example of a **naïve Bayes** model

$\mathbf{P}(\text{Cause}, \text{Effect}_1, \ldots, \text{Effect}_n) = \mathbf{P}(\text{Cause})\, \pi_i \mathbf{P}(\text{Effect}_i \mid \text{Cause})$




Total number of parameters is *linear* in *n*

# Bayesian Networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions.

Syntax:

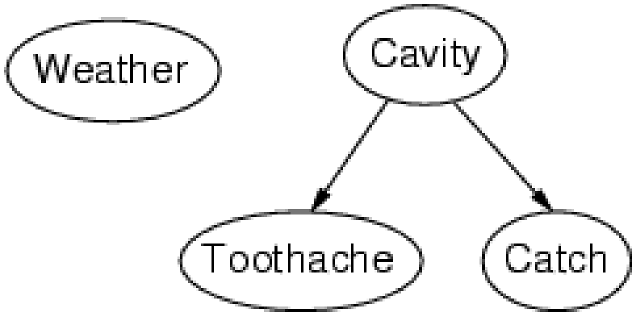
- a set of nodes, one per variable
- a directed, acyclic graph (link ≈ "directly influences")
- a conditional distribution for each node given its parents:

$$\mathbf{P}(X_i \mid Parents(X_i))$$

In the simplest case, conditional distribution represented as a *conditional probability table* (CPT) giving the distribution over $X_i$ for each combination of parent values.

# Example

Topology of network encodes conditional independence assertions:




*Weather* is independent of the other variables

*Toothache* and *Catch* are conditionally independent given *Cavity*

# Example

*I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?*

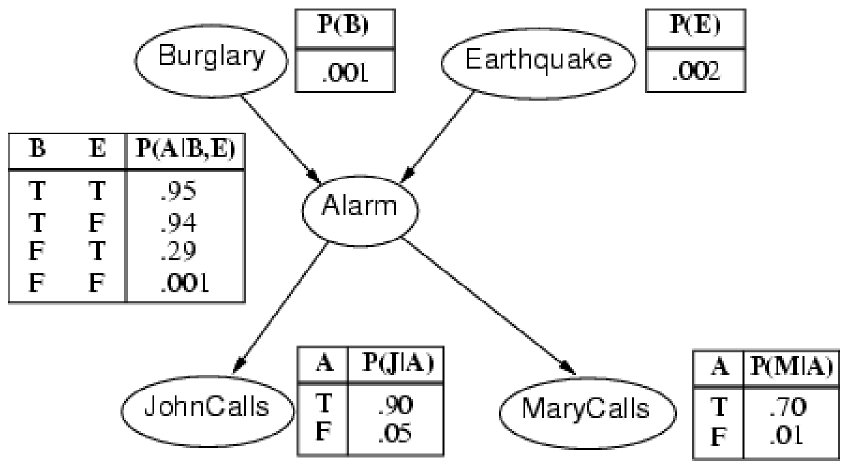
Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

Network topology reflects "causal" knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Notes: Assuming that John and Mary don’t perceive burglary directly; they do not feel minor earthquakes.

# Example contd.

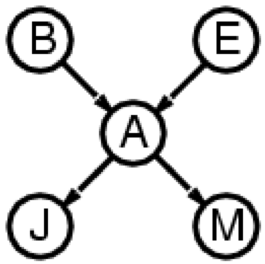



Semantics of Bayesian Networks

– A (more compact) representation of the joint probability distribution; helpful in understanding how to construct network.

– Encoding a collection of conditional independence statements; helpful in understanding how to design inference procedures.

# Compactness




B – Burglary;
E – Earthquake;
A – Alarm;
J – JohnCalls;
M – MaryCalls;

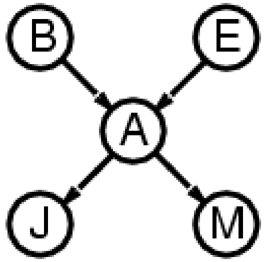
A CPT for Boolean $X_i$ with $k$ Boolean parents has $2^k$ rows for the combinations of parent values.

Each row requires one number $p$ for $X_i = true$
(the number for $X_i = false$ is just $1-p$)

If each variable has no more than $k$ parents, the complete network requires $O(n \cdot 2k)$ numbers.
i.e., grows linearly with $n$, vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5-1 = 31$)

# Global Semantics




B – Burglary;
E – Earthquake;
A – Alarm;
J – JohnCalls;
M – MaryCalls;

The full joint distribution is defined as the product of the local conditional distributions:

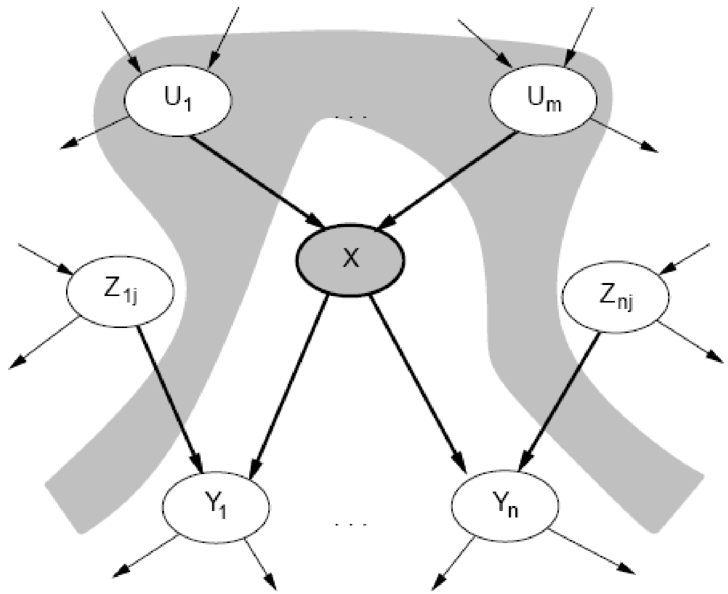
$$P(x_1, x_2, \ldots, x_n) = P(x_1) \times P(x_2 \mid x_1) \times \ldots \times P(x_n \mid x_1 \wedge \ldots \wedge x_{n-1})$$
$$= \Pi_{i=1}^{n} P(x_i \mid x_1 \wedge \ldots \wedge x_{i-1})$$
$$= \Pi_{i=1}^{n} P(x_i \mid Parent(X_i))$$

e.g., $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = \mathbf{P}(J \mid A)\, \mathbf{P}(M \mid A)\, \mathbf{P}(A \mid \neg B, \neg E)\, \mathbf{P}(\neg B)\, \mathbf{P}(\neg E)$
$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

# Local Semantics

Local semantics: each node is conditionally independent of its nondescendants given its parents

(e.g. *JohnCalls* is independent of *Burglary* and *Earthquake*, given the value of *Alarm*).

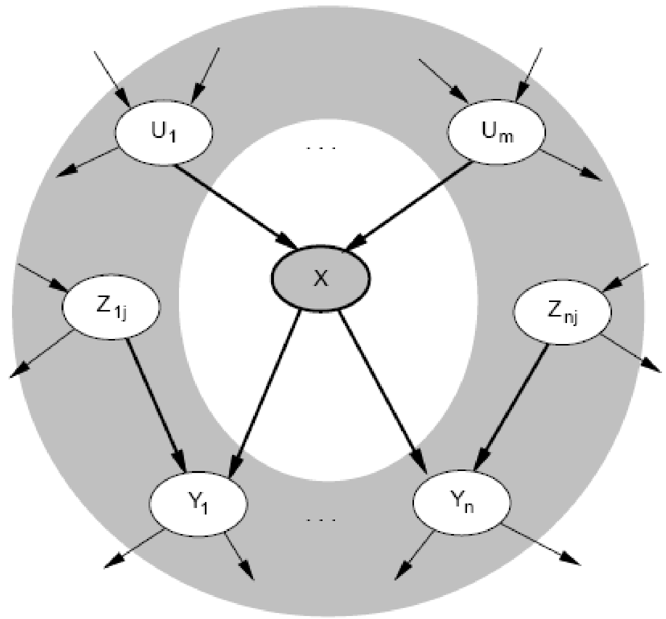



Theorem: Local semantics ⇔ global semantics

# Markov Blanket

Each node is conditionally independent of all others given its **Markov blanket**:

*parents + children + children's parents*




E.g. *Burglary* is independent of *JohnCalls* and *MaryCalls*, given *Alarm* and *Earthquake*.

# Constructing Bayesian Networks

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

1. Choose a set of relevant variables $X_i$ that describe the domain.

2. Choose an ordering for the variables.

3. While there are variables left:
   a) Pick a variable $X_i$ and add a node to the network for it.
   b) Set Parent($X_i$) to some minimal set of nodes already in the net such that the conditional independence property is satisfied:

$$P(x_i \mid x_1, \ldots, x_{i-1}) = P(x_i \mid Parent(X_i))$$

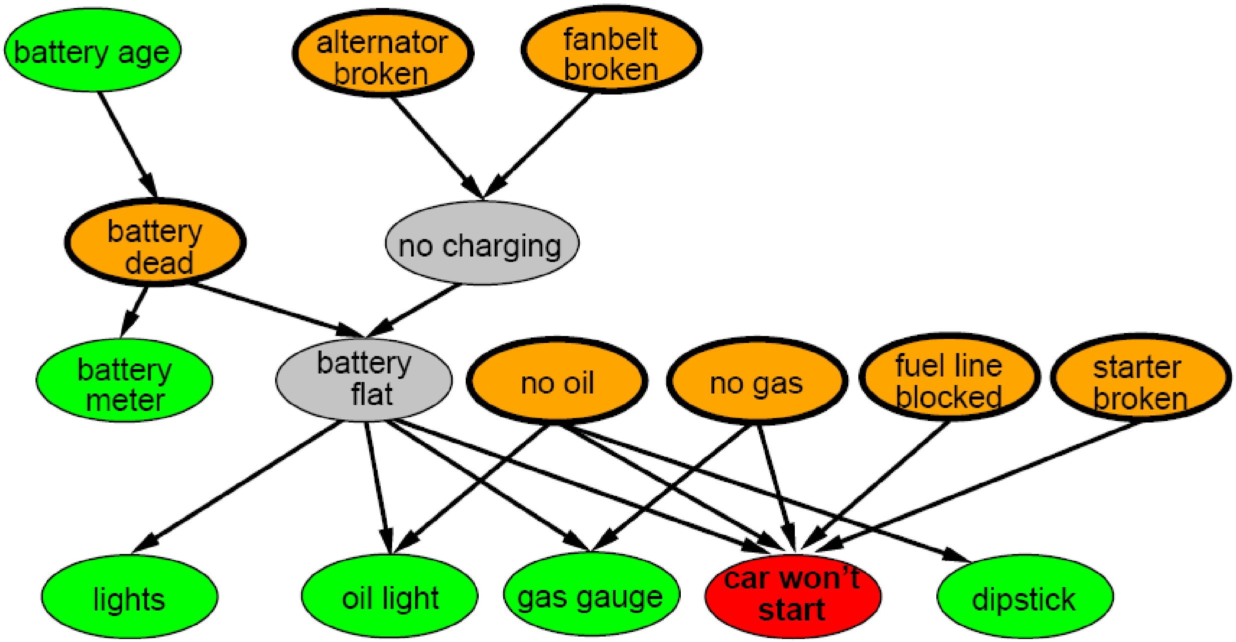
   c) Define the CPT for $X_i$.

# Example: Car Diagnosis

Initial evidence: car won't start

Testable variables (green), "broken, so fix it" variables (orange)

Hidden variables (gray)



# Compactness and Node Ordering

Compactness of Bayesian Network is an example of a locally structured (or sparse) system.

The correct order to add nodes is to add the “root causes” first, then the variable they influence, so on until “leaves” reached.

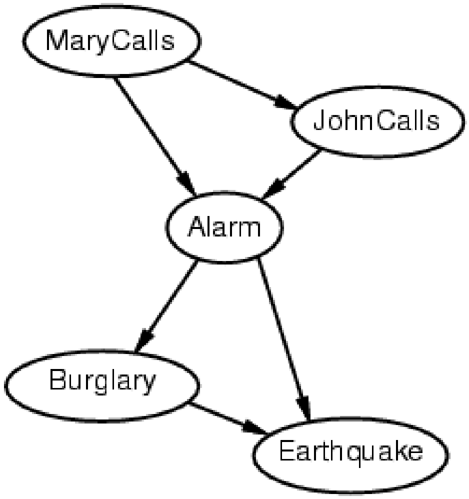
An example of wrong ordering

- Suppose we choose the ordering M, J, A, B, E

- The network is less compact:

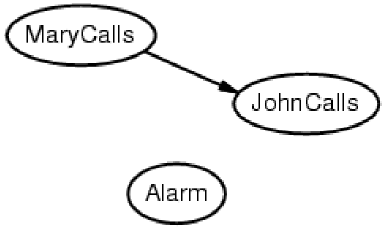
1 + 2 + 4 + 2 + 4 = 13 numbers needed

See the procedure of construction next.



# Example

Suppose we choose the ordering M, J, A, B, E

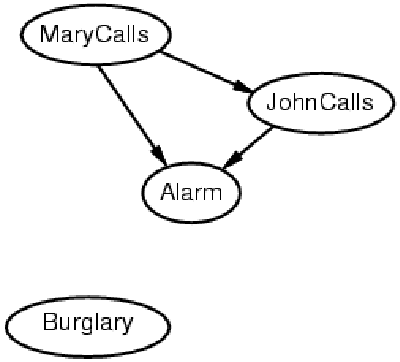



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

# Example…

Suppose we choose the ordering M, J, A, B, E




$P(J \mid M) = P(J)$? **No**

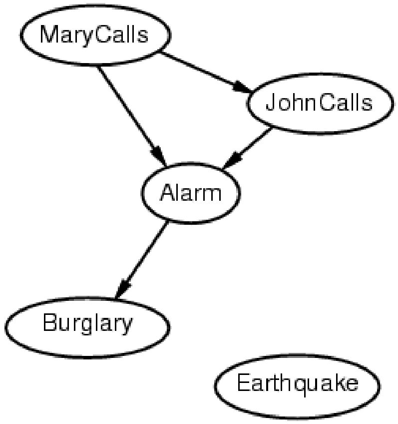
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? **No**

$P(B \mid A, J, M) = P(B \mid A)$?

$P(B \mid A, J, M) = P(B)$?

# Example...

Suppose we choose the ordering M, J, A, B, E




$P(J \mid M) = P(J)$? **No**

$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? **No**

$P(B \mid A, J, M) = P(B \mid A)$? **Yes**

$P(B \mid A, J, M) = P(B)$? **No**

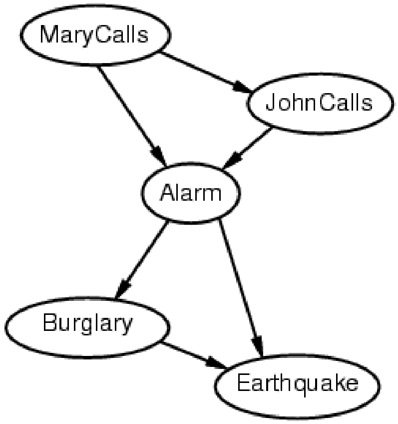
$P(E \mid B, A, J, M) = P(E \mid A)$?

$P(E \mid B, A, J, M) = P(E \mid A, B)$?

# Example…

Suppose we choose the ordering M, J, A, B, E

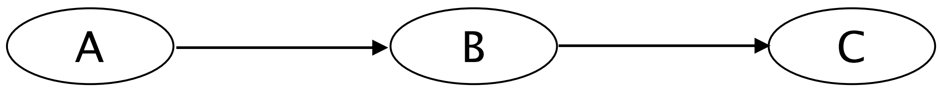
$P(J \mid M) = P(J)$ ? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$ ? No
$P(B \mid A, J, M) = P(B \mid A)$ ? Yes
$P(B \mid A, J, M) = P(B)$ ? No
$P(E \mid B, A ,J, M) = P(E \mid A)$ ? No
$P(E \mid B, A, J, M) = P(E \mid A, B)$ ? Yes




- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

# Conditional Independence: Causal Chains

Causal chains give rise to conditional independence:



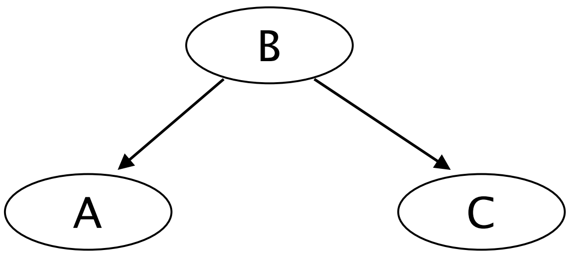

$$P(C|A , B) = P(C|B)$$

Example:

- A = excessive exercises
- B = dehydrated
- C = hard to concentrate

## Conditional Independence: Common Causes

Common causes (or ancestors) give rise to conditional independence:



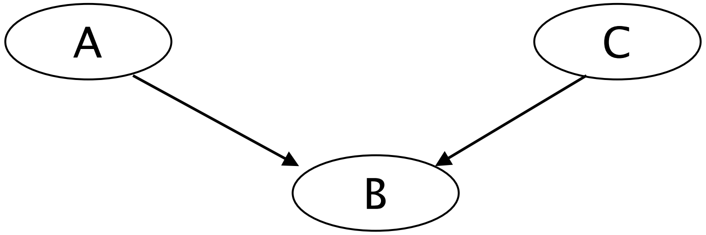

$$P(C|A, B) = P(C|B)$$

Example:

- A = thirsty
- B = dehydrated
- C = hard to concentrate

# Conditional Independence: Common Effects

Common effects (or descendents) give rise to conditional dependence:




$$P(C|A, B) \neq P(C|B)$$

Example:

– A = excessive exercises
– B = dehydrated
– C = inadequate water intake

# Summary

- Probability is a rigorous formalism for uncertain knowledge.
- Joint probability distribution specifies probability of every atomic event.
- Queries can be answered by summing over atomic events.
- For non-trivial domains, we must find a way to reduce the joint size.
- Independence and conditional independence provide the tools.
- Bayes nets provide a natural representation for (causally induced) conditional independence.
- Topology + CPTs = compact representation of joint distribution.
- Generally easy for domain experts to construct.

**Acknowledgement**: the slides were developed based on notes from Russell & Norvig's text, and several RMIT computer science staff members over the years.