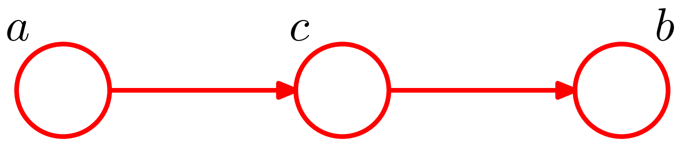
Announcements

Assignment three posted. It is due Friday, Feb 24 (truthfully Feb 28).

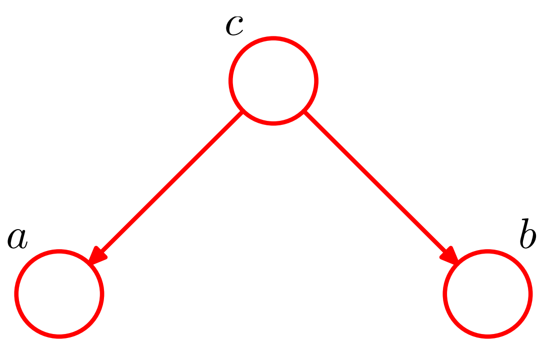
Office hours in GS 919 at 10:30 to 11:30.

Back to case one



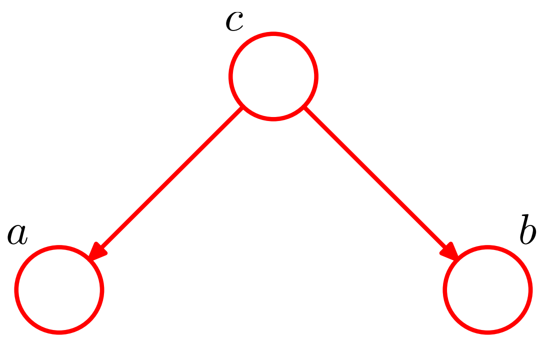
- Let a="smokes", c="high blood pressure", b="stroke"
- p(c|a) tells you probability of having high blood pressure if you smoke (for some definition of each).

Can we distinguish case two from case one?



- Let a=“smokes”, c=“high blood pressure”, b=“stroke”
- p(a|c) tells you probability of being a smoker if you have high blood pressure (for some definition of each).

Can we distinguish case two from case one?



- Let a=“smokes”, b=“high blood pressure”, c=“stroke”
- p(a|c) tells you probability of being a smoker if you have high blood pressure (for some definition of each).
- Data for estimating p(c|a) in first case, and p(a|c) in second case cannot tell you which model you should prefer.
  - “Correlation is not causation”
- Causality implied by our generative process is about the statistics of the data, not physical causality.

More on causality

- References
  - Kollar and Friedman, Chapter 21 which starts on page 1009!
  - Classic book by Pearl, Causality: Models, Reasoning, and Inference, 2000
    - A version is available on-line (bayes.cs.ucla.edu/BOOK-99/book-toc.html)

More on causality

- We have been focussed on the joint distribution which is adequate (arguably optimal) for answering the queries we have studied
- In particular, we know how distributions over unknowns change due to evidence
- For many problems (e.g., computer vision and much of machine learning) this is sufficient
  - Either causes are obvious or not relevant

More on causality

- Two correlated variables can have multiple equivalent graphs hinting at **different** causal stores able to provide the **same** joint.
  - A causes B
  - B causes A
  - C causes both A and B
  - A and B cause C (and A and B are correlated by explaining away)

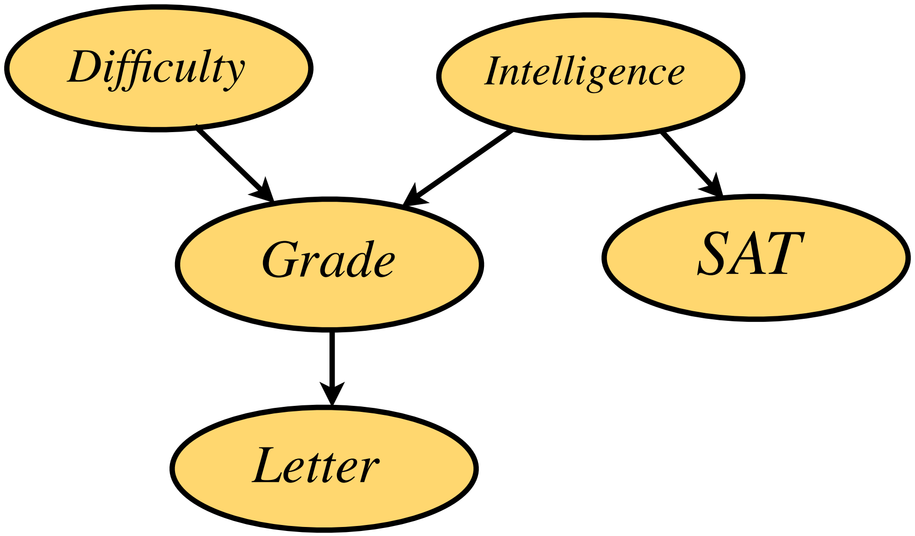
- Given a choice, we prefer the Bayes net that also represents our causal theory (if we have one)
  - More natural, easier to understand
  - Helps tell you whether observed statistics are consistent with your theory
    - (Covered briefly next)

Intervention

- Two Bayes nets that give the same joint distribution can differ in what they say about an intervention.
- We represent an intervention, x, as setting some subset of the variables, X, to the value, x, denoted by $do(X=x)$.
  - Example 1: Creating an experimental group that will not smoke
  - Example 2: Setting your grade to A by hacking into a computer
- On the surface, this might look like conditioning on X, but it is different --- the graph needs to change also
  - We need to "mutilate" the graph

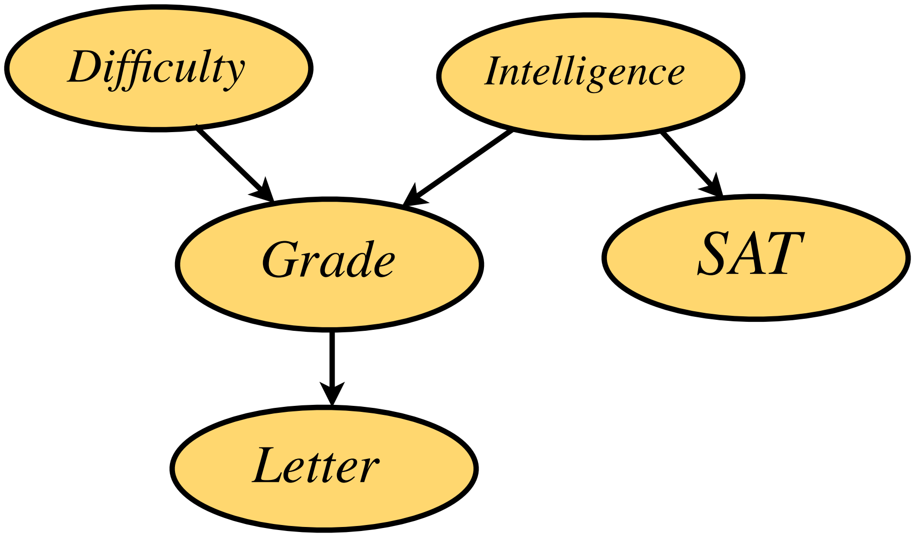
Representing Intervention

- Example one (students and grades, again)
  - Does observing grade change your belief about SAT?



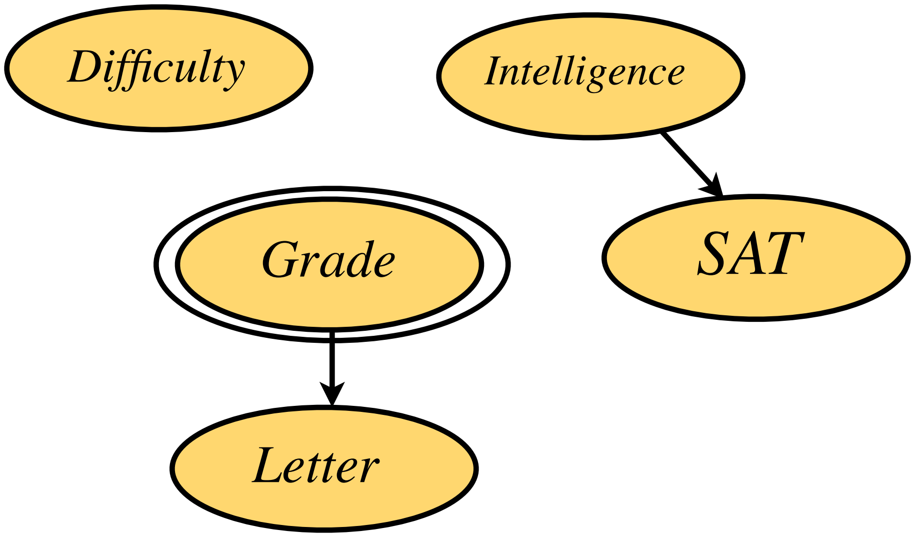
Representing Intervention

- Example one (students and grades, again)
  - Does observing grade change your belief about SAT?
- Now, suppose we intervene on the *Grade* random variable
  - E.G., we fix it by hacking into the grade computer
  - Now does observing grade change your belief about SAT?



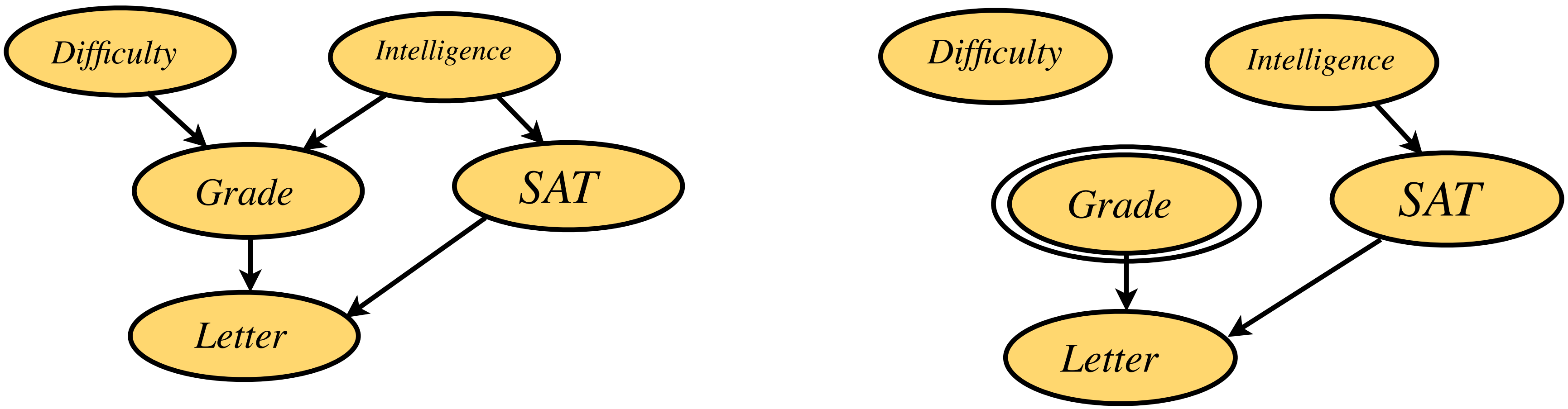
Representing Intervention

- The intervention not only conditions on the variable, it cuts the links that influence it. This is the mutilated graph.



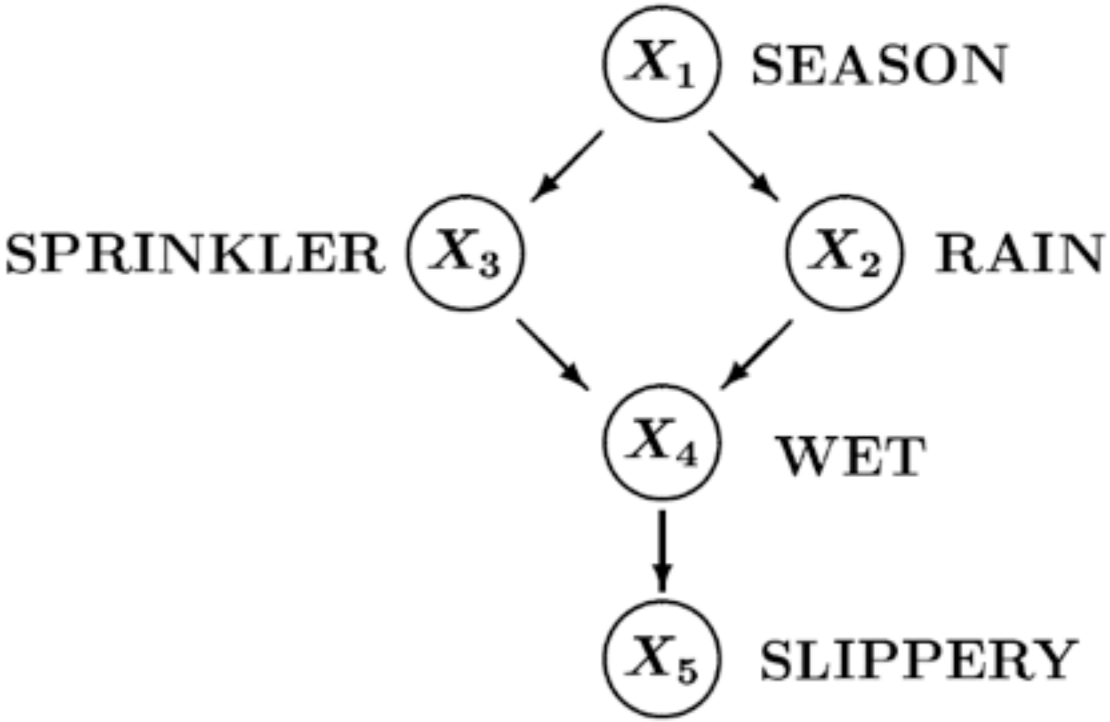
Representing Intervention

- Another example --- the student from before with a link between SAT and letter. Now we expect that the intervention does not entirely explain the letter, but that the influence of *grade* is direct (only).



Representing Intervention

- Another example --- from Pearl, 2000.
  - Consider the intervention of turning the sprinkler “on”



Representing Intervention

- Representation of the intervention of turning the sprinkler on.

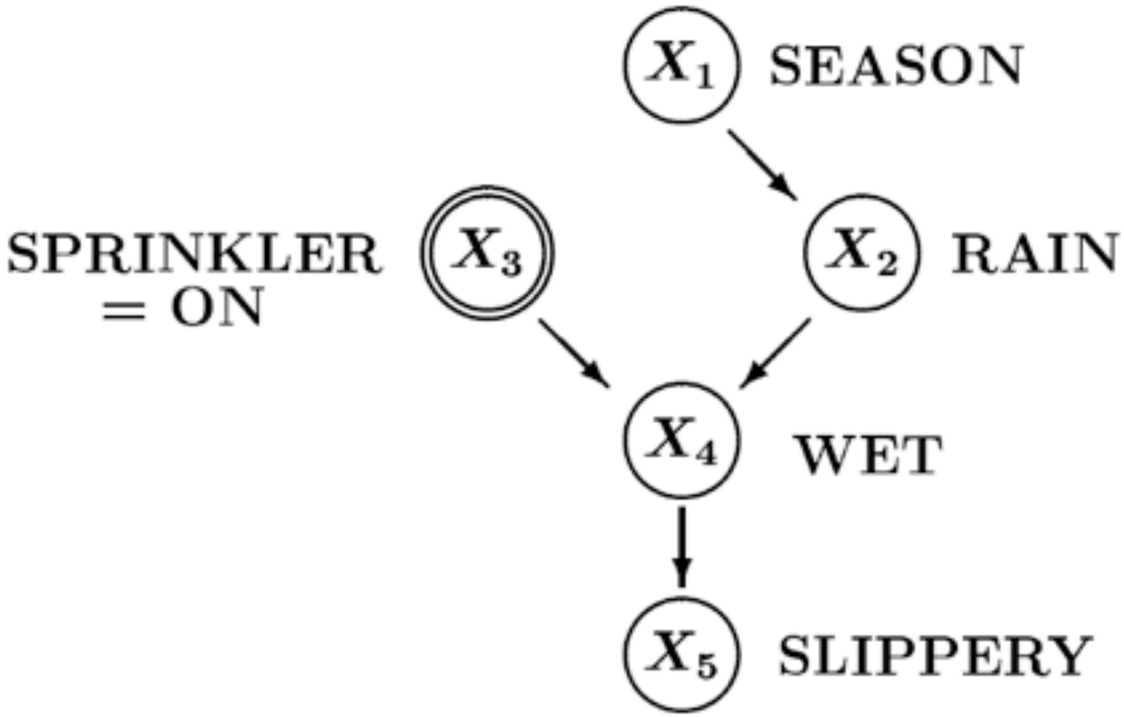


Figure 1.4: Network representation of the action "turning the sprinkler On."

Back to graphs in general

Can graphs capture all independence?

- Do our graphs faithfully capture the independence structure of our distributions?
- Recall that

G is an I-map for P if $I(G) \subseteq I(P)$

In other words, all independence represented in G are true.

(There could be more independence in P that G does not reveal).

- Hence we are asking if $I(G) \equiv I(P)$

Since $I(G) \subseteq I(P)$ this amounts to asking if $I(P) \subseteq I(G)$

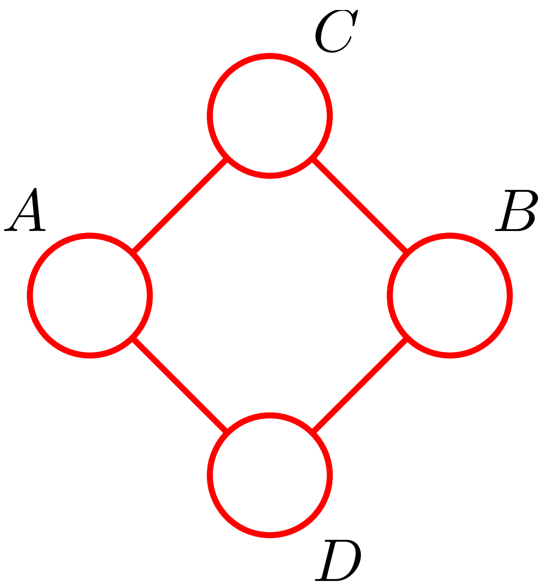
Perfection

G is an P-map for P if $I(G) \equiv I(P)$ (perfect map)

In other words, all independence represented in G are true, and there are no other independence relations.

Do all distributions have perfect maps?

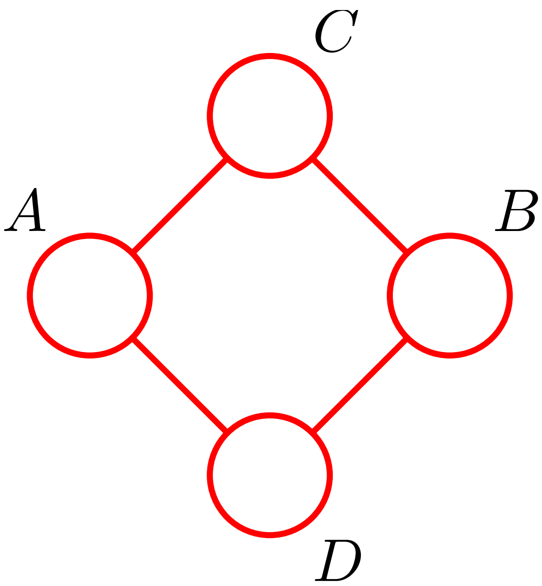
Perfection may not be attainable



The "misconception" example in K&F (pp. 82-3), where Alice, Bob, Charles, and Debbie study in pairs shown, but A and B never work together, nor do C and D.

Note **no arrows**, but a link still means some probabilistic relation.

Perfection may not be attainable



Suppose that we have

$$\left(A \perp B \middle| C, D\right)$$

and

$$\left(C \perp D \middle| A, B\right)$$

Now, draw the Bayes net (have fun!).

Note **no arrows**, but a link still means some probabilistic relation.

Interesting questions

- Does every probability distribution have a corresponding Bayesian network?

Chain rule says yes

- Given the independence structure of a probability distribution, and a graph that captures them all (I(G)=I(P), is the corresponding graph unique (ignoring isomorphisms)?

Case study of three nodes says no

- Do our graphs **always** faithfully capture the independence structure of our distributions?

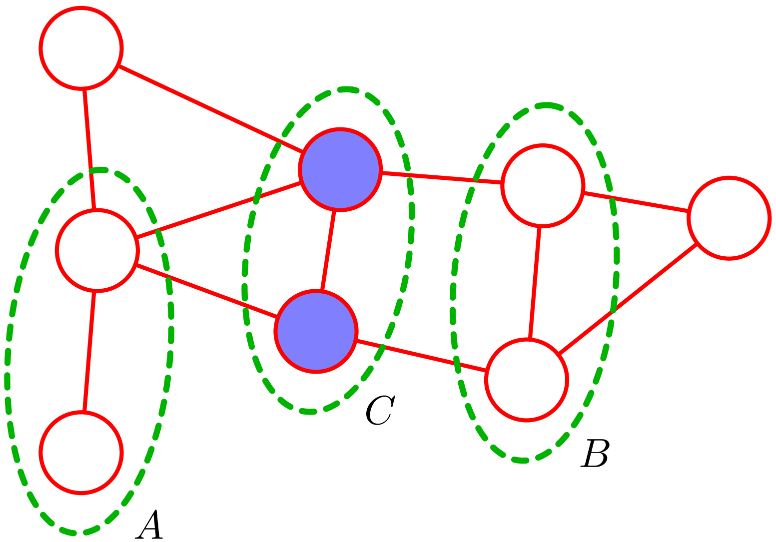
Misconception example says no

Undirected graphical models

- Also referred to as
  - Markov Networks
  - Markov Random Fields
- Nodes represent (groups of) random variables
- Edges represent probabilistic relations between connected nodes.
- We have already seen an example suggestive that arrows are not always helpful.

Undirected graphical models

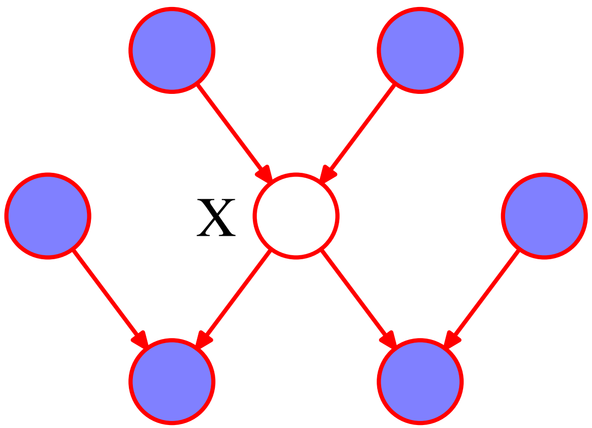
- The analog to d-separation is simper
  - Disjoint sets A and B are independent conditioned on C if all paths from nodes in A to nodes in B pass through C.

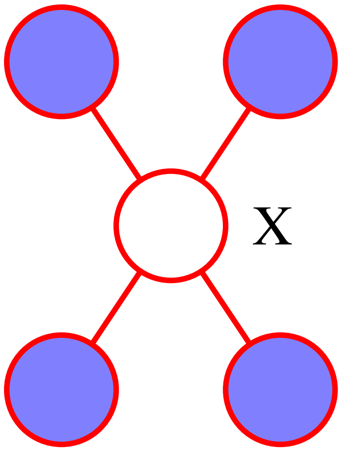


Here $\left(A \perp B | C\right)$ for all probability distributions represented by this graph.

Markov Blanket

- The Markov blanket of a node, X, is a particular set of (nearby) nodes B where $X \perp X_i \mid B$ for all X_i
- For directed graphs the Markov blanket is the parents, children, and co-parents of X.
- For undirected graphs this is simply the set of nodes connected to X.





Undirected graphical models

- Bayes nets where nodes only have one parent are easily converted to undirected graphs without changing links.

- (Discussed in more detail soon)