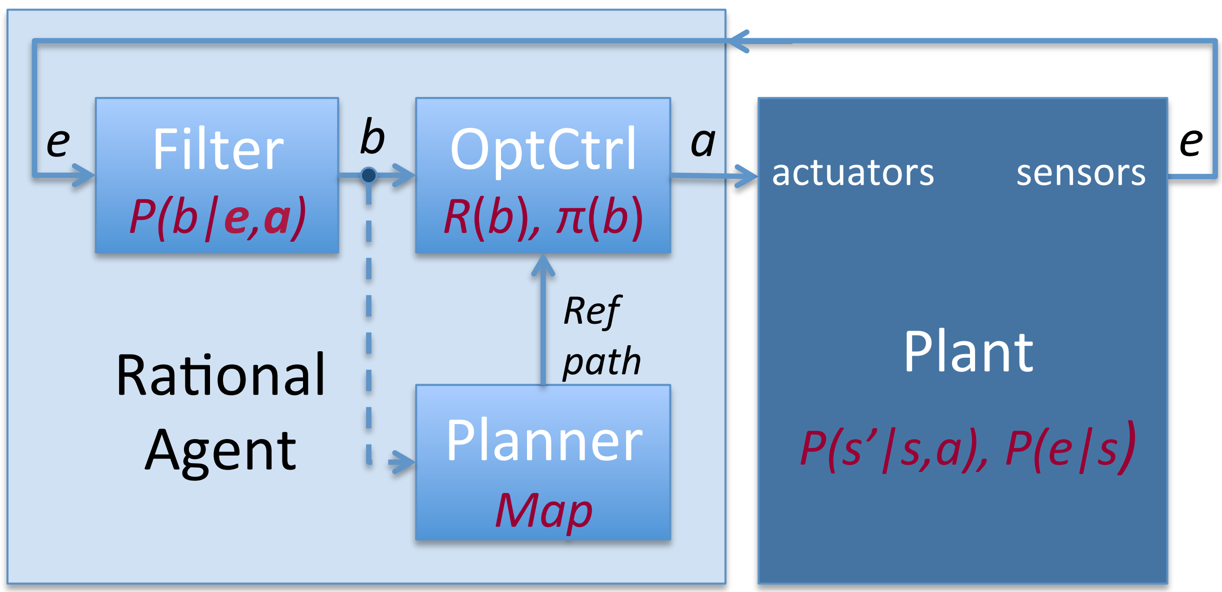
Probabilistic Reasoning

Chapter 14 (Models)



Probabilistic Models

- Models describe how (a portion of) the world works

Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly. For n variables with domain size d, joint table has d^n entries --- exponential in n.

Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly. For n variables with domain size d, joint table has d^n entries --- exponential in n.

  - Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly. For n variables with domain size d, joint table has d^n entries --- exponential in n.
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets
- Conditional Independences

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference

Bayes' Nets

- Representation
  - Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

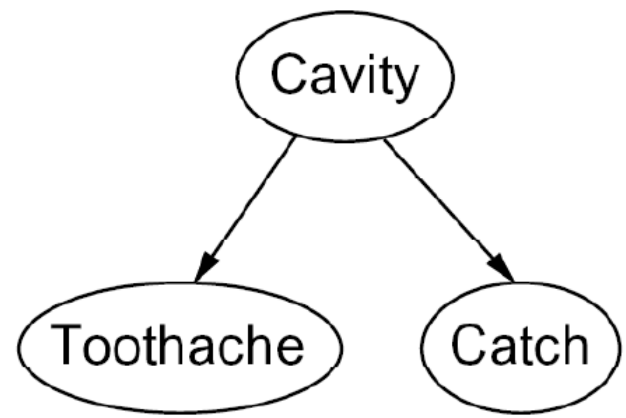
Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

Graphical Model Notation

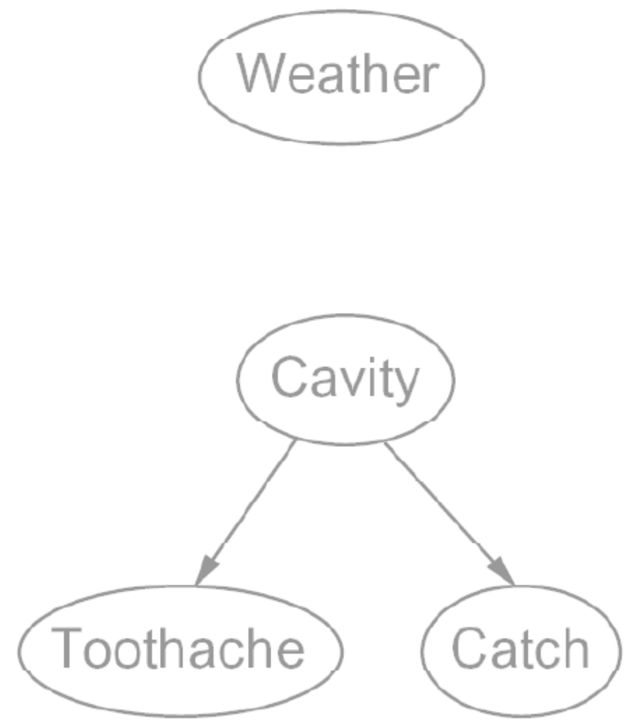
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)



Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don’t!)



Example: Coin Flips

- N independent coin flips

Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

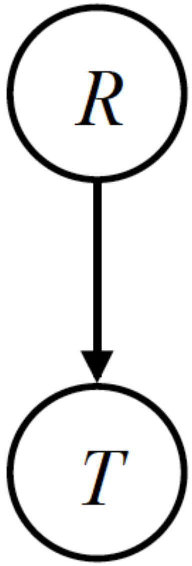
Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

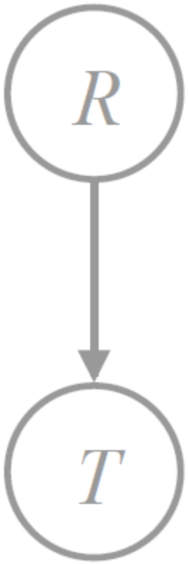
Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: independence
- Model 2: rain causes traffic



Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: independence
- Model 2: rain causes traffic
- Why is an agent using model 2 better?



Example: Traffic II

- Let's build a causal graphical model

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net

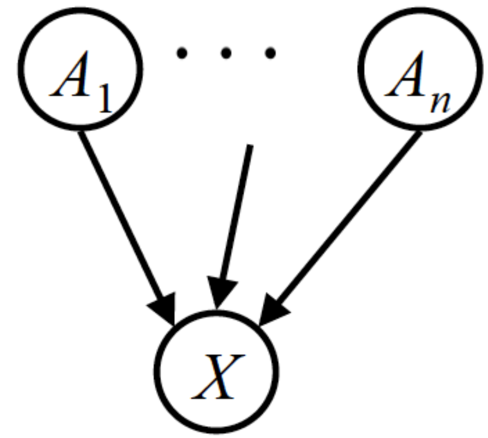
Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net

- A set of nodes, one per variable X

Bayes' Net Semantics

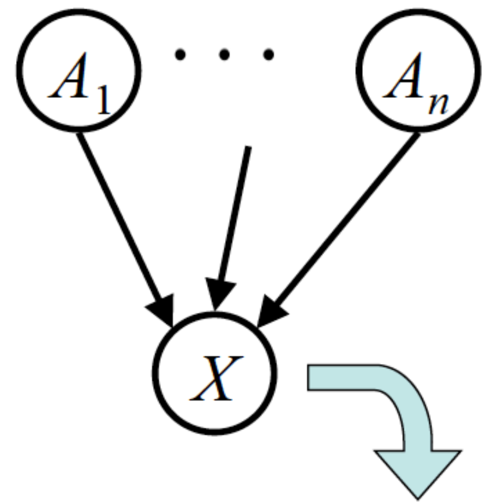
- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph



Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \ldots a_n)$$



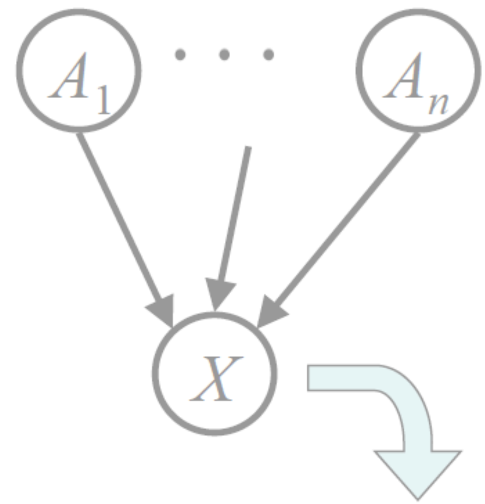
$$P(X|A_1 \ldots A_n)$$

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

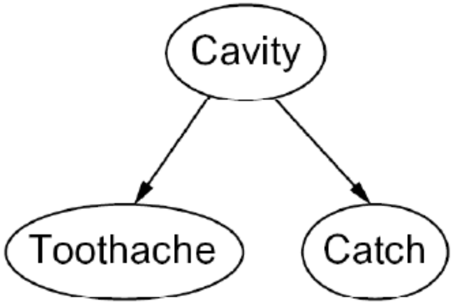
$$P(X|a_1 \dots a_n)$$

  - CPT: conditional probability table
  - Description of a noisy "causal" process



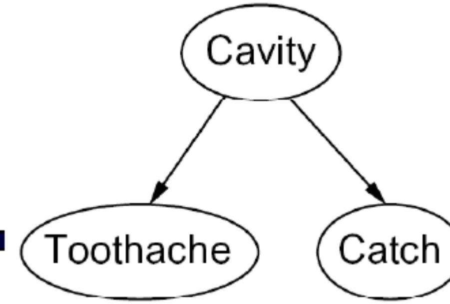
$$P(X|A_1 \dots A_n)$$

Probabilities in BNs



- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions

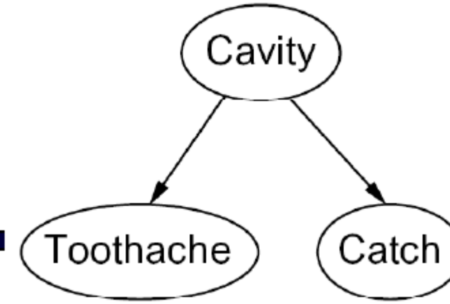
Probabilities in BNs



- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

Probabilities in BNs



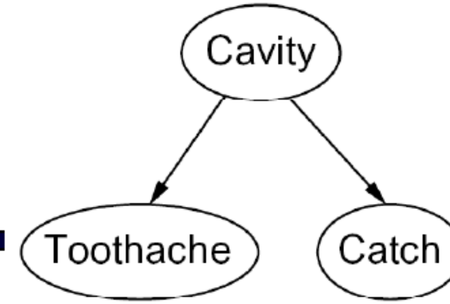
- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Example:

$$P(+cavity, +catch, \neg toothache)$$

Probabilities in BNs



- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

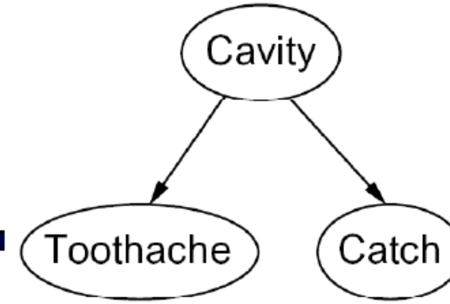
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Example:

$$P(+cavity, +catch, \neg toothache)$$

- This lets us reconstruct any entry of the full joint

Probabilities in BNs



- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Example:

$$P(+cavity, +catch, \neg toothache)$$

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

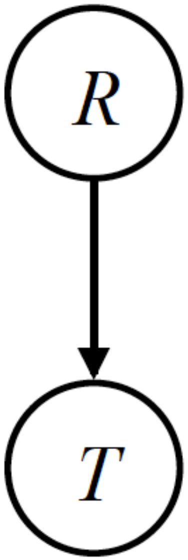
$P(X_n)$

h	0.5
t	0.5

$$P(h, h, t, h) =$$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

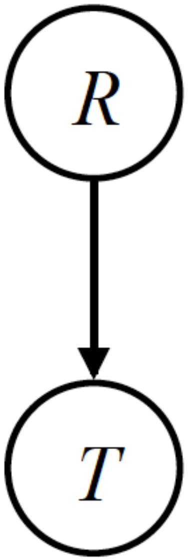
Example: Traffic



$P(R)$

+r	1/4
¬r	3/4

Example: Traffic



$P(R)$

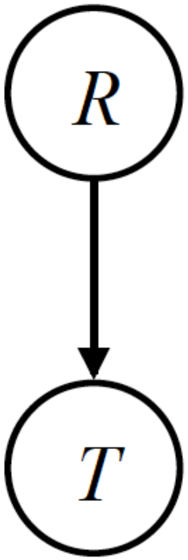
+r	1/4
¬r	3/4

$P(T|R)$

+r

+t	3/4
¬t	1/4

Example: Traffic



$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

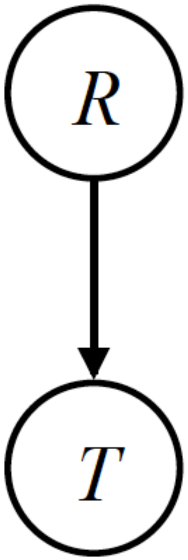
+r →

+t	3/4
¬t	1/4

¬r →

+t	1/2
¬t	1/2

Example: Traffic



$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

+r →

+t	3/4
¬t	1/4

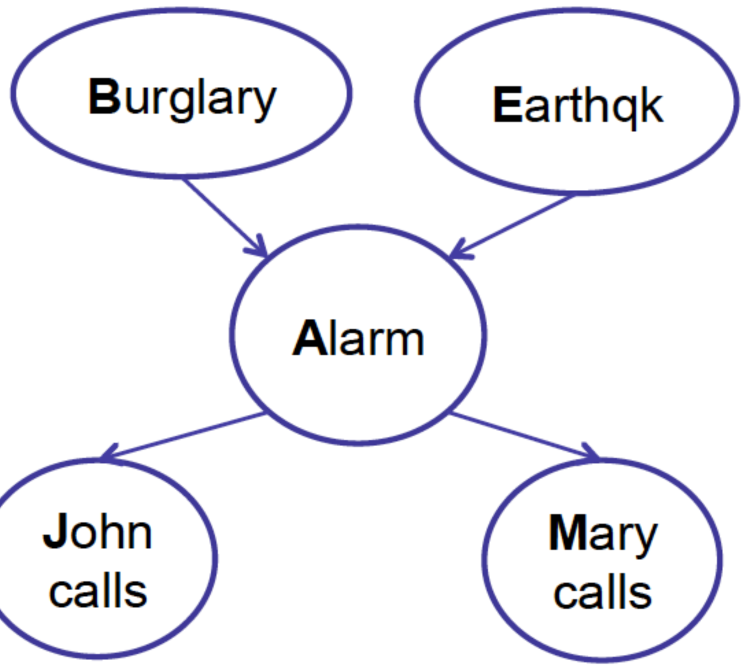
¬r →

+t	1/2
¬t	1/2

$$P(+r, \neg t) =$$

Example: Alarm Network

B	P(B)
+b	0.001
¬b	0.999



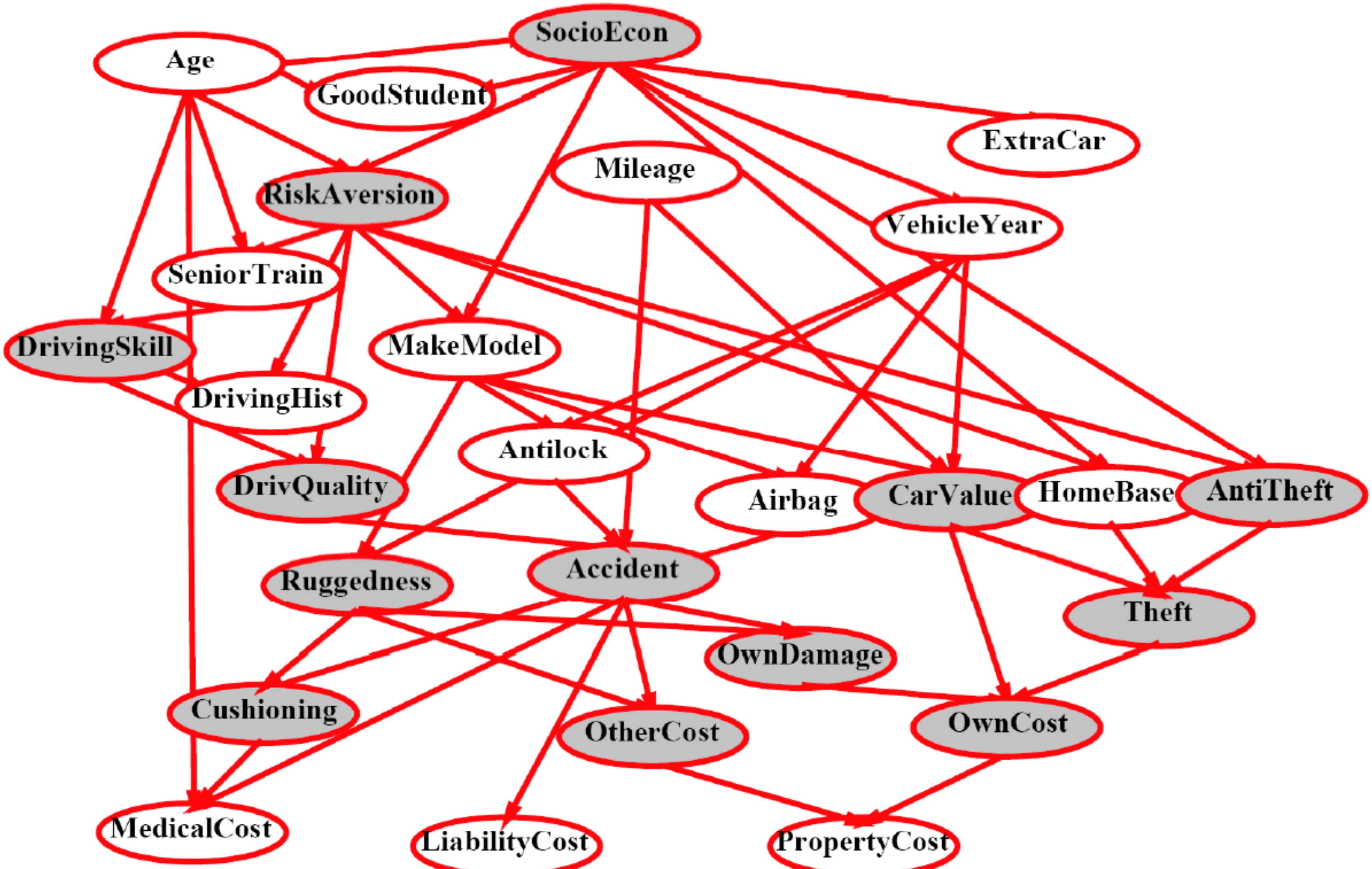
E	P(E)
+e	0.002
¬e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

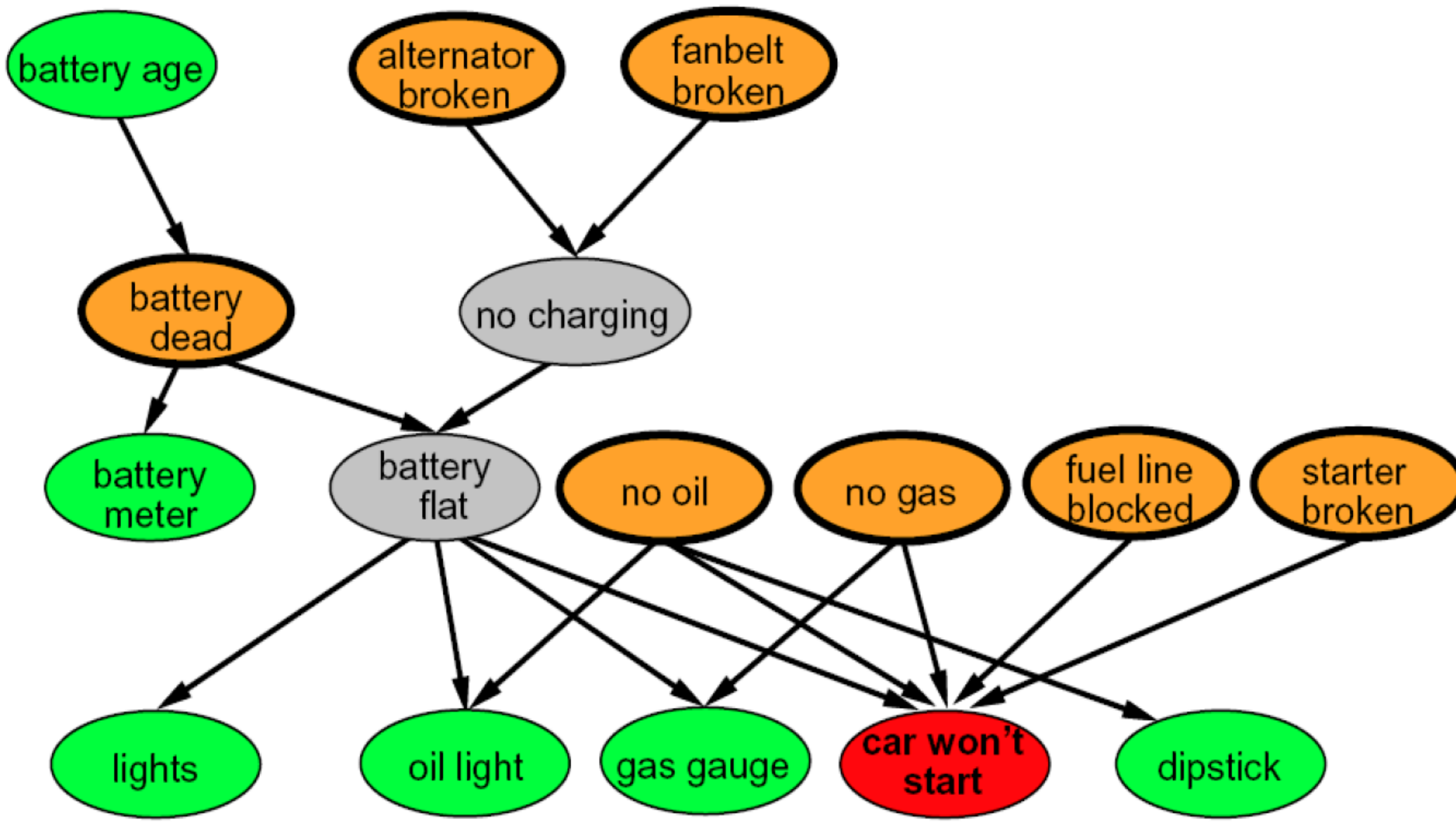
A	J	P(J|A)
+a	+j	0.9
+a	¬j	0.1
¬a	+j	0.05
¬a	¬j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	¬m	0.3
¬a	+m	0.01
¬a	¬m	0.99

Example Bayes' Net: Insurance



Example Bayes' Net: Car



Build your own Bayes nets!

- http://www.aispace.org/bayes/index.shtml

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

Size of a Bayes’ Net

- How big is a joint distribution over N Boolean variables?

  2^N

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$
- BNs: Huge space savings!

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

- How big is an N-node net if nodes have up to k parents?

  $O(N * 2^{k+1})$

- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Bayes' Nets

- Representation
  - ✓ Informal first introduction of Bayes' nets through causality "intuition"
  - More formal introduction of Bayes' nets

Representing Joint Probability Distributions

- **Table representation**:

  number of parameters: d^n-1

Representing Joint Probability Distributions

- **Table representation**:

  number of parameters: d^n-1

- **Chain rule representation:**

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$

  number of parameters: $(d-1) + d(d-1) + d^2(d-1)+\ldots+d^{n-1}(d-1) = d^n-1$

  Size of CPT = (number of different joint instantiations of the preceding variables) *times* (number of values current variable can take on *minus* 1)

Representing Joint Probability Distributions

- **Table representation**:

  number of parameters: d^n-1

- **Chain rule representation:**

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$

  number of parameters: $(d-1) + d(d-1) + d^2(d-1)+\ldots+d^{n-1}(d-1) = d^n-1$

  Size of CPT = (number of different joint instantiations of the preceding variables) *times* (number of values current variable can take on *minus* 1)

- Both can represent any distribution over the n random variables. Makes sense same number of parameters needs to be stored.

Representing Joint Probability Distributions

- **Table representation**:

  number of parameters: d^n-1

- **Chain rule representation:**

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$

  number of parameters: $(d-1) + d(d-1) + d^2(d-1)+\ldots+d^{n-1}(d-1) = d^n-1$

  Size of CPT = (number of different joint instantiations of the preceding variables) *times* (number of values current variable can take on *minus* 1)

- Both can represent any distribution over the n random variables. Makes sense same number of parameters needs to be stored.
- Chain rule applies to all orderings of the variables, so for a given distribution we can represent it in n! = n factorial = n(n-1)(n-2)…2.1 different ways with the chain rule

Chain Rule → Bayes' net

- **Chain rule representation: applies to ALL distributions**
  - Pick any ordering of variables, rename accordingly as $x_1, x_2, \ldots, x_n$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i | x_1 \ldots x_{i-1})$$

Exponential in n

number of parameters: $(d-1) + d(d-1) + d^2(d-1) + \ldots + d^{n-1}(d-1) = d^n - 1$

Chain Rule → Bayes' net

- **Chain rule representation: applies to ALL distributions**
  - Pick any ordering of variables, rename accordingly as $x_1, x_2, \ldots, x_n$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i | x_1 \ldots x_{i-1})$$

Exponential in n

number of parameters: $(d-1) + d(d-1) + d^2(d-1) + \ldots + d^{n-1}(d-1) = d^n - 1$

- **Bayes' net representation: makes assumptions**
  - Pick any ordering of variables, rename accordingly as $x_1, x_2, \ldots, x_n$
  - Pick any directed acyclic graph consistent with the ordering
  - *Assume* following conditional independencies:

$$P(x_i | x_1 \cdots x_{i-1}) = P(x_i | parents(X_i))$$

→ **Joint:** $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$

number of parameters: (maximum number of parents = K)

$$\sum_{i=1}^{n} d^{|\text{parents}(X_i)|}(d-1) = O(nd^K(d-1)) = O(nd^{K+1})$$

Linear in n

Note: no causality assumption made anywhere.

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

Causality?

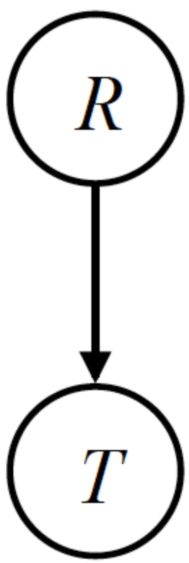
- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology only guaranteed to encode conditional independence

Example: Traffic

- Basic traffic net
- Let’ s multiply out the joint



$P(R)$

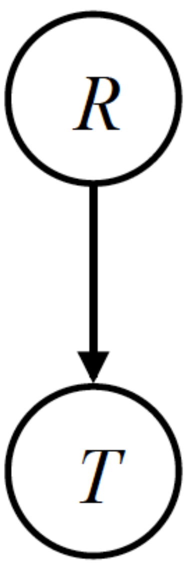
r	1/4
¬r	3/4

$P(T|R)$

r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

Example: Traffic

- Basic traffic net
- Let's multiply out the joint



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

r	t	3/4
	¬t	1/4

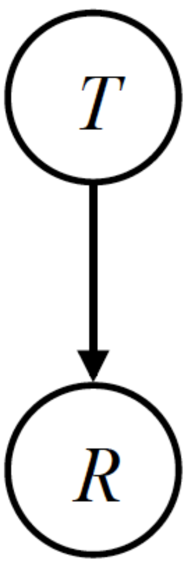
¬r	t	1/2
	¬t	1/2

$P(T, R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Example: Reverse Traffic

- Reverse causality?



$P(T)$

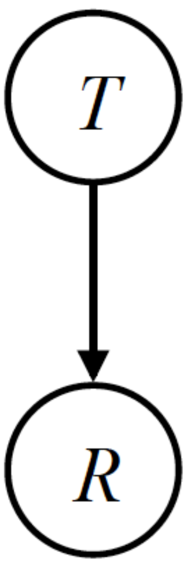
t	9/16
¬t	7/16

$P(R|T)$

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

Example: Reverse Traffic

- Reverse causality?



$P(T)$

t	9/16
¬t	7/16

$P(R|T)$

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

$P(T, R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence

$P(X_1)$

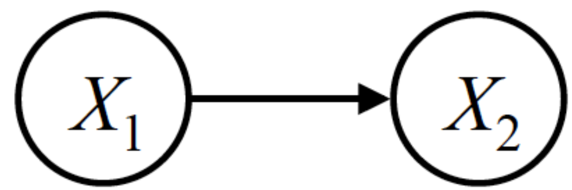
h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence



$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$P(X_1)$

h	0.5
t	0.5

$P(X_2|X_1)$

h | h	0.5
t | h	0.5

h | t	0.5
t | t	0.5

Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$P(X_1)$

h	0.5
t	0.5

$P(X_2|X_1)$

h | h	0.5
t | h	0.5

h | t	0.5
t | t	0.5

- Adding unneeded arcs isn't wrong, it's just inefficient

Bayes' Nets

- Representation
  - ✓ Informal first introduction of Bayes' nets through causality "intuition"
  - ✓ More formal introduction of Bayes' nets
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

Bayes Nets: Assumptions

- To go from chain rule to Bayes' net representation, we made the following assumption about the distribution:

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

Bayes Nets: Assumptions

- To go from chain rule to Bayes' net representation, we made the following assumption about the distribution:

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Turns out that probability distributions that satisfy the above ("chain-rule→Bayes net") conditional independence assumptions
  - often can be guaranteed to have many more conditional independences
  - These guaranteed additional conditional independences can be read off directly from the graph

Bayes Nets: Assumptions

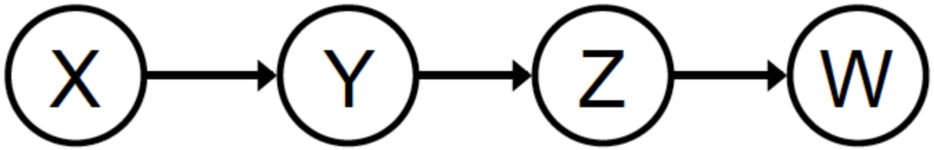
- To go from chain rule to Bayes' net representation, we made the following assumption about the distribution:

$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Turns out that probability distributions that satisfy the above ("chain-rule→Bayes net") conditional independence assumptions
  - often can be guaranteed to have many more conditional independences
  - These guaranteed additional conditional independences can be read off directly from the graph

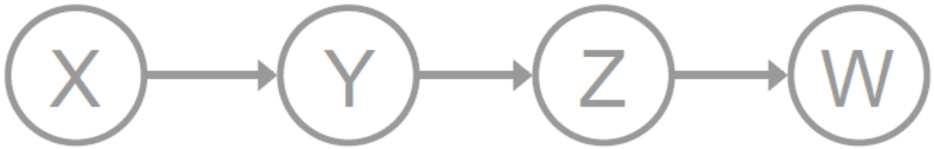
- Important for modeling: understand assumptions made when choosing a Bayes net graph

Example



- Conditional independence assumptions directly from simplifications in chain rule:

Example



- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?

Independence in a BN

- Given a Bayes net graph
  - Important question:

    Are two nodes guaranteed to be independent given certain evidence?

Independence in a BN

- Given a Bayes net graph
  - Important question:

    Are two nodes guaranteed to be independent given certain evidence?

  Equivalent question:

    Are two nodes independent given the evidence in all distributions that can be encoded with the Bayes net graph?

Independence in a BN

- Given a Bayes net graph
  - Important question:

    Are two nodes guaranteed to be independent given certain evidence?

  Equivalent question:

    Are two nodes independent given the evidence in all distributions that can be encoded with the Bayes net graph?

- Before proceeding: How about opposite question: Are two nodes guaranteed to be *dependent* given certain evidence?

Independence in a BN

- Given a Bayes net graph
  - Important question:

    Are two nodes guaranteed to be independent given certain evidence?

  Equivalent question:

    Are two nodes independent given the evidence in all distributions that can be encoded with the Bayes net graph?

- Before proceeding: How about opposite question: Are two nodes guaranteed to be *dependent* given certain evidence?
  - No! For any BN graph you can choose all CPT's such that all variables are independent by having P(X | Pa(X) = paX) not depend on the value of the parents. Simple way of doing so: pick all entries in all CPTs equal to 0.5 (assuming binary variables)

D-separation: Outline

- Study independence properties for triples

D-separation: Outline

- Study independence properties for triples

- Any complex example can be analyzed by considering relevant triples

Causal Chains

- This configuration is a "causal chain"

$$X \rightarrow Y \rightarrow Z$$

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z ?

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x,y,z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z ? *No!*

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

  - Is it guaranteed that X is independent of Z ? *No!*
    - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z ? *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example: $P(y|x) = 1$ if $y=x$, 0 otherwise
    $P(z|y) = 1$ if $z=y$, 0 otherwise

    Then we have $P(z|x) = 1$ if $z=x$, 0 otherwise

    hence X and Z are not independent in this example

Causal Chains

- This configuration is a “causal chain”

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

$$= P(z|y)$$

Yes!

Causal Chains

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Is it guaranteed that X is independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

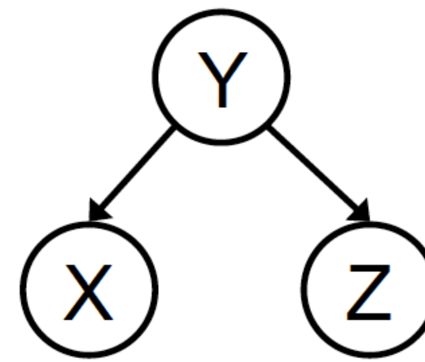
$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence

Common Cause

- Another basic configuration: two effects of the same cause
  - Is it guaranteed that X and Z are independent?



Y: Project due

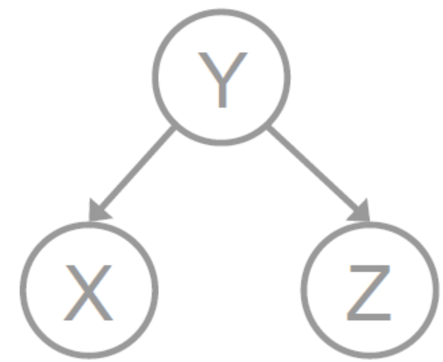
X: Piazza busy

Z: Lab full

Common Cause

- Another basic configuration: two effects of the same cause
  - Is it guaranteed that X and Z are independent?
    - *No!*
    - Counterexample:

    Choose $P(X|Y)=1$ if $x=y$, 0 otherwise,
    Choose $P(z|y) = 1$ if $z=y$, 0 otherwise.



Y: Project due

X: Piazza busy

Z: Lab full

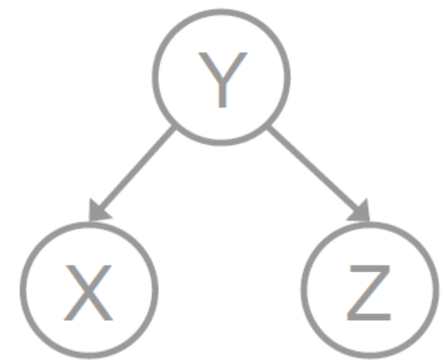
Common Cause

- Another basic configuration: two effects of the same cause
  - Is it guaranteed that X and Z are independent?
    - *No!*
    - Counterexample:

    Choose P(X|Y)=1 if x=y, 0 otherwise,

    Choose P(z|y) = 1 if z=y, 0 otherwise.

    Then P(x|z)=1 if x=z and 0 otherwise, hence X and Z are not independent in this example and hence it is not guaranteed that if a distribution can be encoded with the Bayes' net structure on the right that X and Z are independent in that distribution



Y: Project due

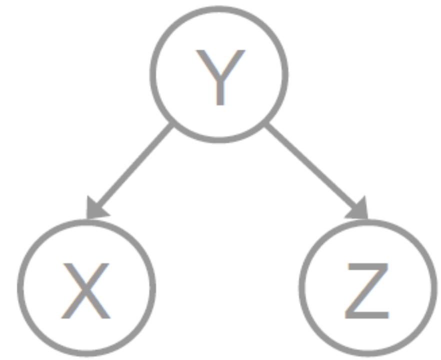
X: Piazza busy

Z: Lab full

Common Cause

- Another basic configuration: two effects of the same cause

  - Is it guaranteed that X and Z are independent given Y?



Y: Project due

X: Piazza busy

Z: Lab full

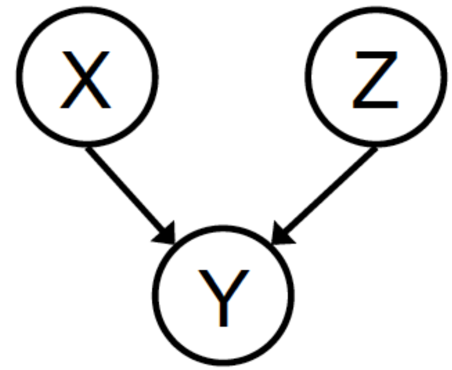
$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$ *Yes!*

- Observing the cause blocks influence between effects.

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?



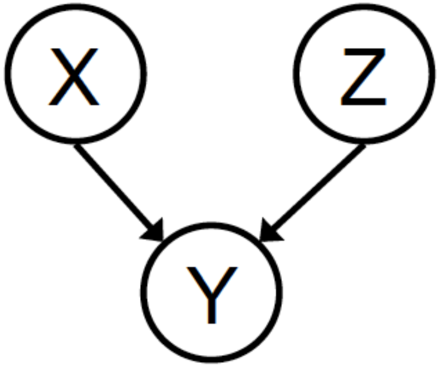
X: Raining

Z: Ballgame

Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - *Yes*: the ballgame and the rain cause traffic, but they are not correlated



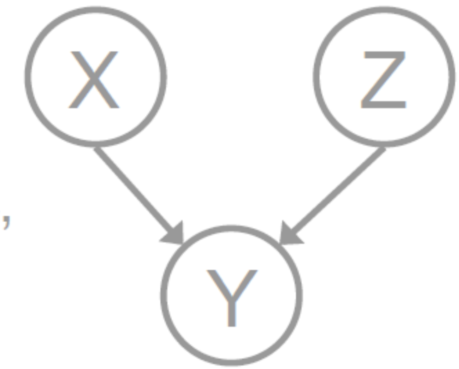
X: Raining

Z: Ballgame

Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)



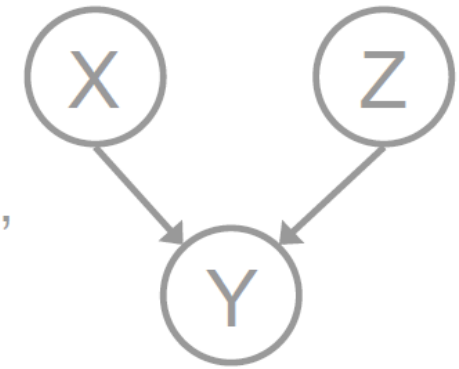
X: Raining

Z: Ballgame

Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?



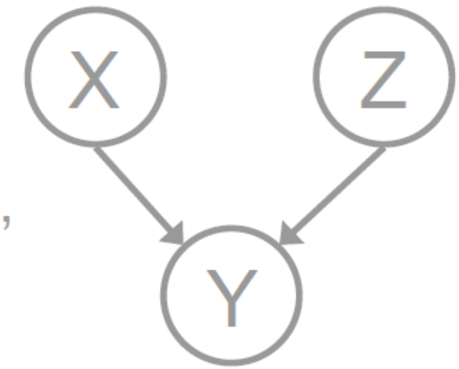
X: Raining

Z: Ballgame

Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - *No*: seeing traffic puts the rain and the ballgame in competition as explanation?



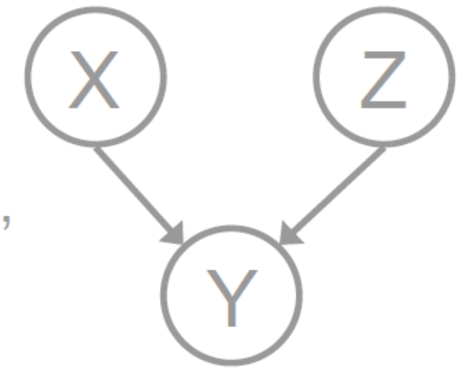
X: Raining

Z: Ballgame

Y: Traffic

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - *No*: seeing traffic puts the rain and the ballgame in competition as explanation?
  - This is backwards from the other cases
    - Observing an effect activates influence between possible causes.



X: Raining

Z: Ballgame

Y: Traffic

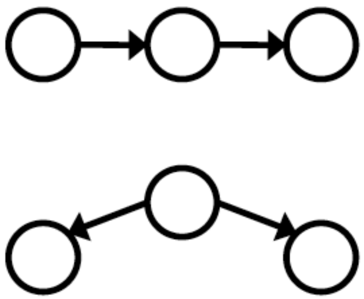
Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?

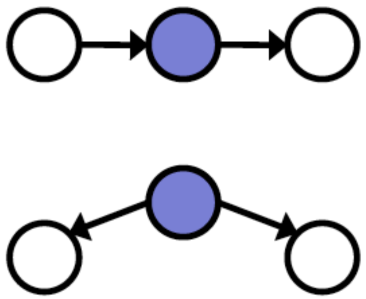
Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!

Active Triples

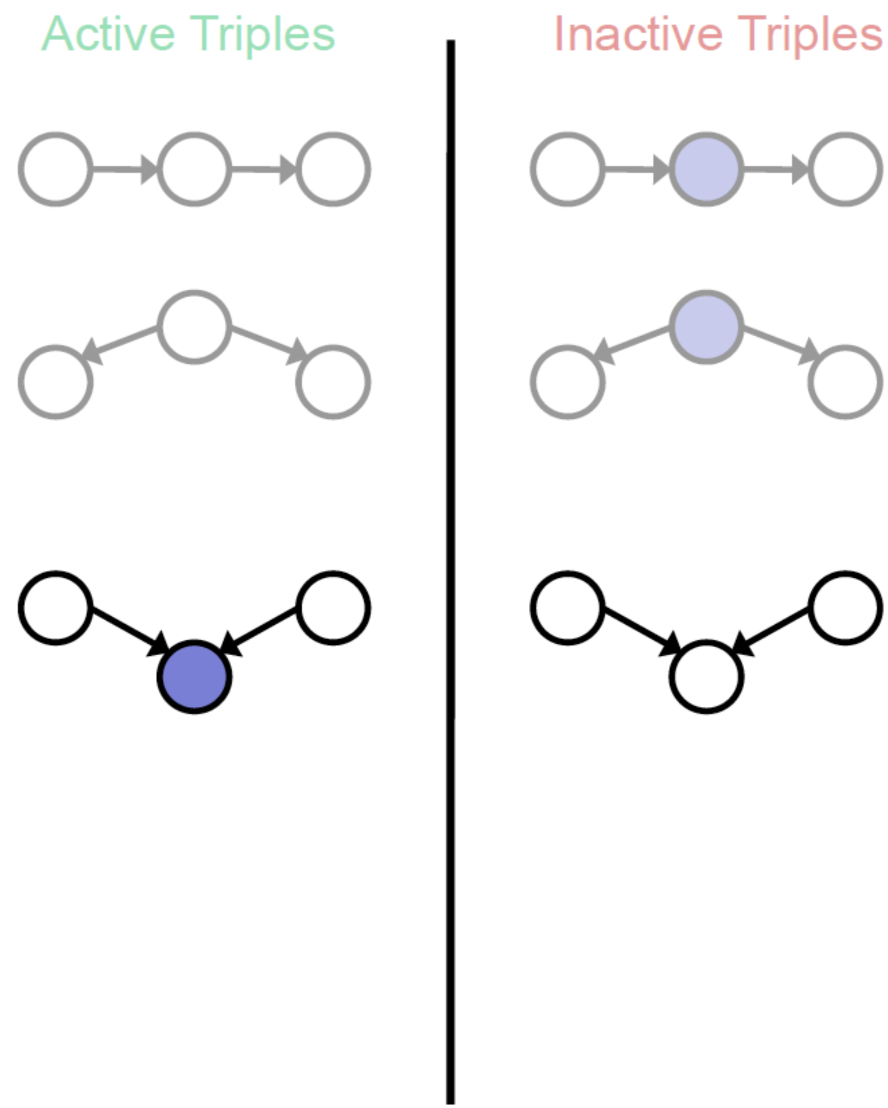


Inactive Triples



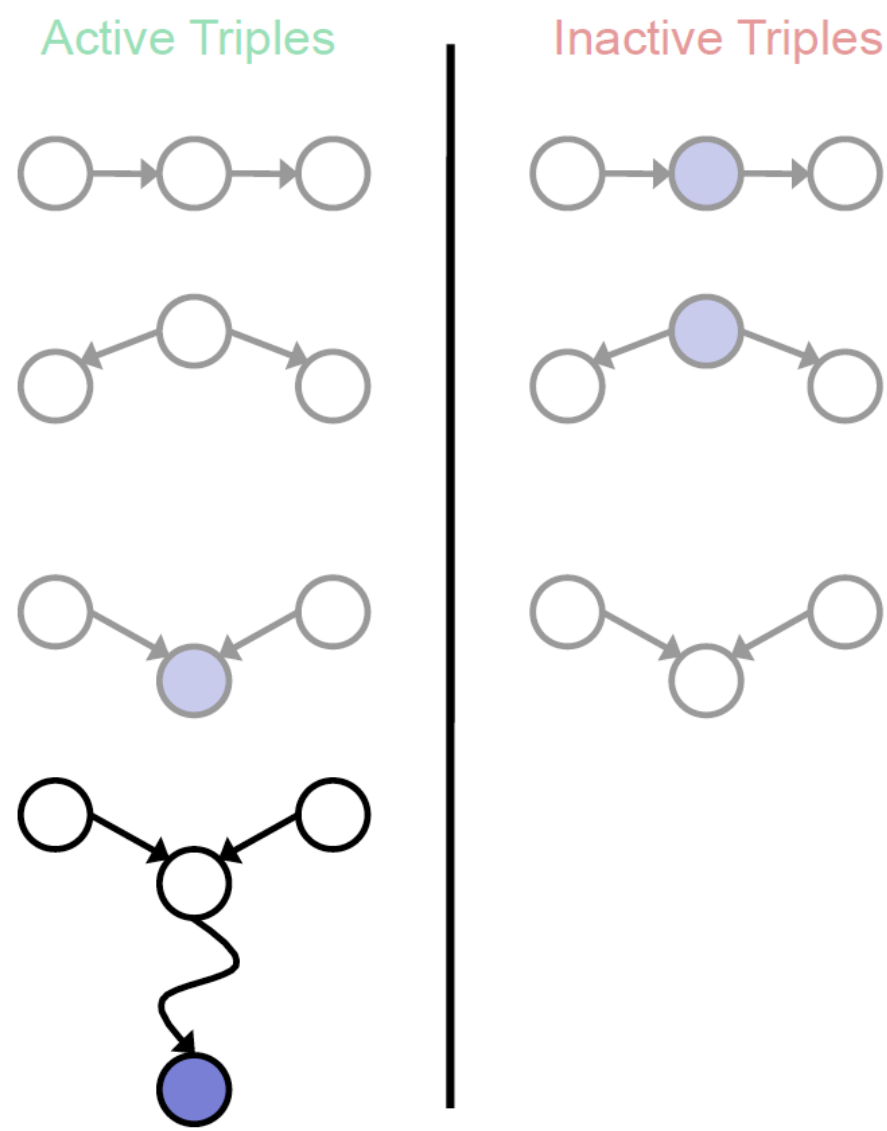
Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain $A \to B \to C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \to C$ where B is unobserved



Reachability (D-Separation)

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
  - Common effect (aka v-structure) $A \rightarrow B \leftarrow C$ where B *or one of its descendents* is observed



D-Separation

- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$

D-Separation

- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Shade all evidence nodes

D-Separation

- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Shade all evidence nodes
- For all (undirected!) paths between and
  - Check whether path is active
    - If active return:

      not guaranteed that $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$

D-Separation

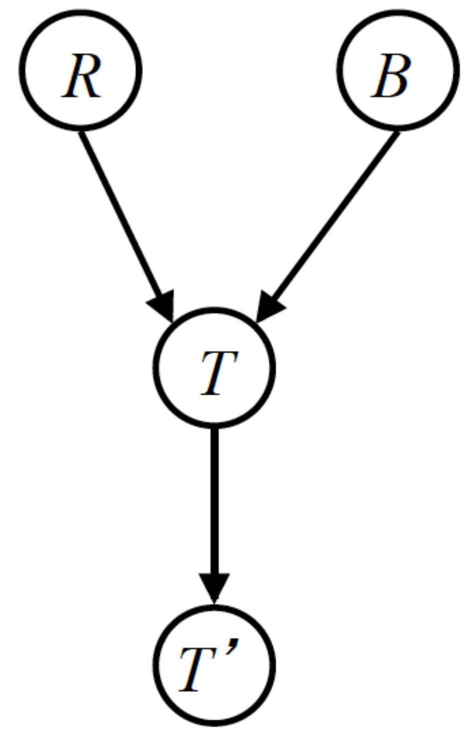
- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Shade all evidence nodes
- For all (undirected!) paths between and
  - Check whether path is active
    - If active return:
      not guaranteed that $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$
- (If reaching this point all paths have been checked and shown inactive)
  - Return: guaranteed tat $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$

Example

$$R \perp\!\!\!\perp B$$

$$R \perp\!\!\!\perp B | T$$

$$R \perp\!\!\!\perp B | T'$$

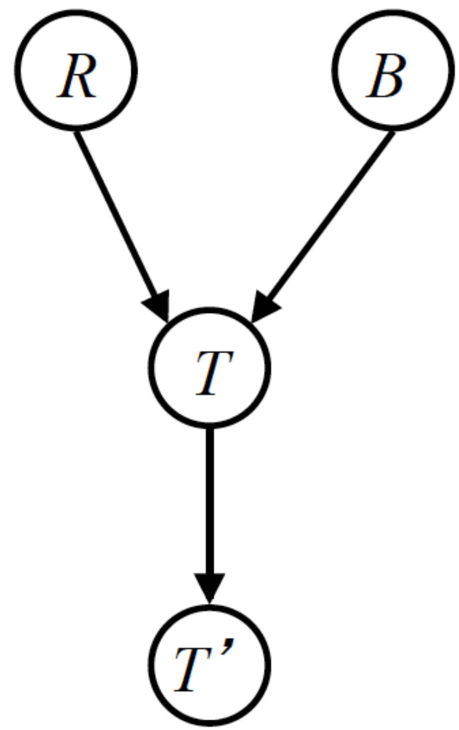


Example

$R \perp\!\!\!\perp B$ Yes

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example

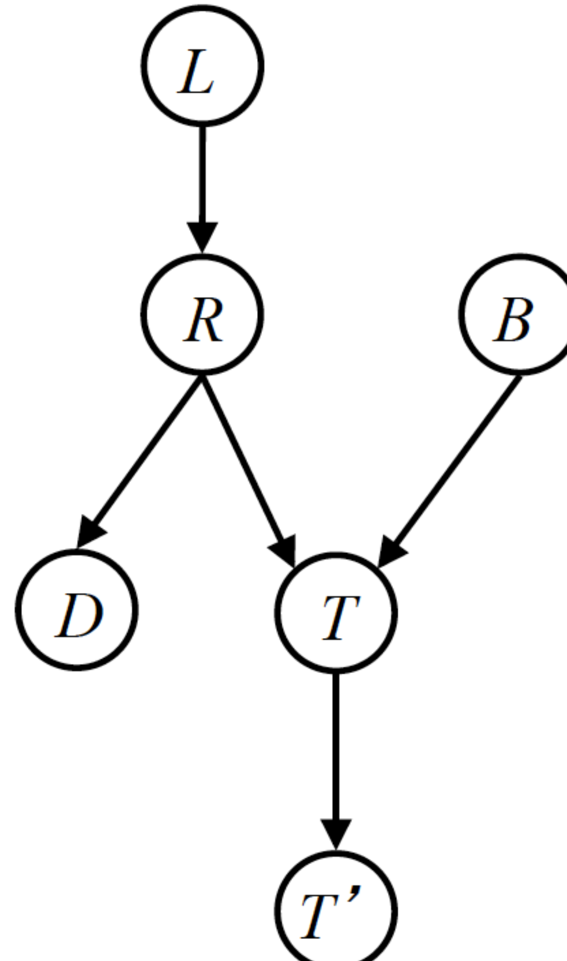
$$L \perp\!\!\!\perp T' | T$$

$$L \perp\!\!\!\perp B$$

$$L \perp\!\!\!\perp B | T$$

$$L \perp\!\!\!\perp B | T'$$

$$L \perp\!\!\!\perp B | T, R$$



Example

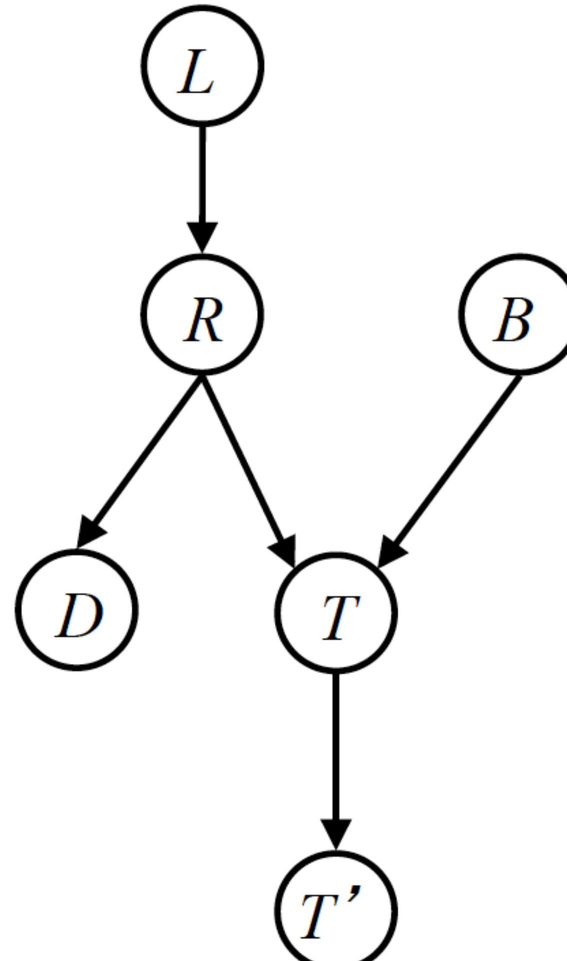
$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

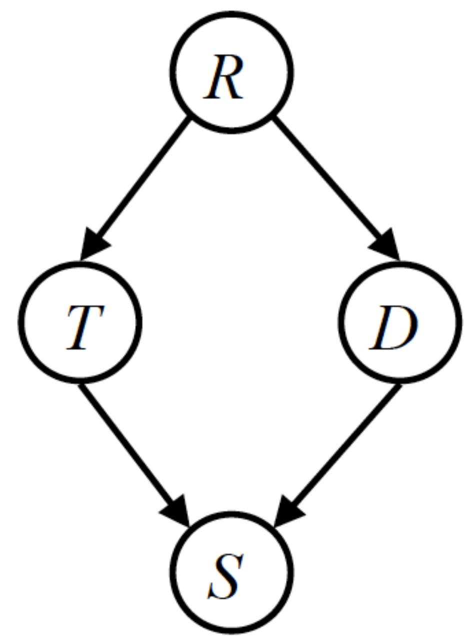
$L \perp\!\!\!\perp B | T, R$ *Yes*



Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I' m sad
- Questions:

$$T \perp\!\!\!\perp D$$

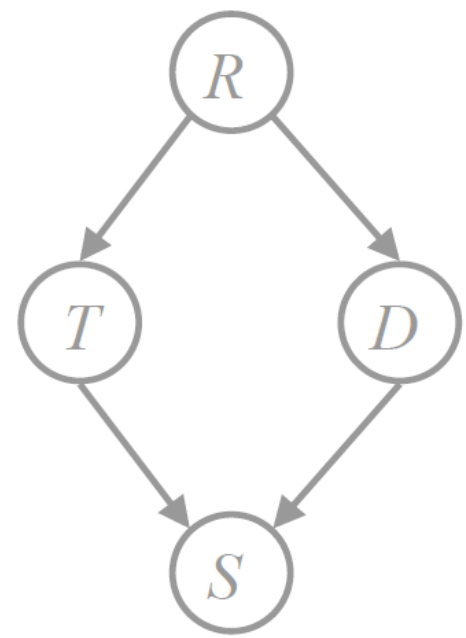


Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I' m sad
- Questions:

$$T \perp\!\!\!\perp D$$

$$T \perp\!\!\!\perp D | R$$

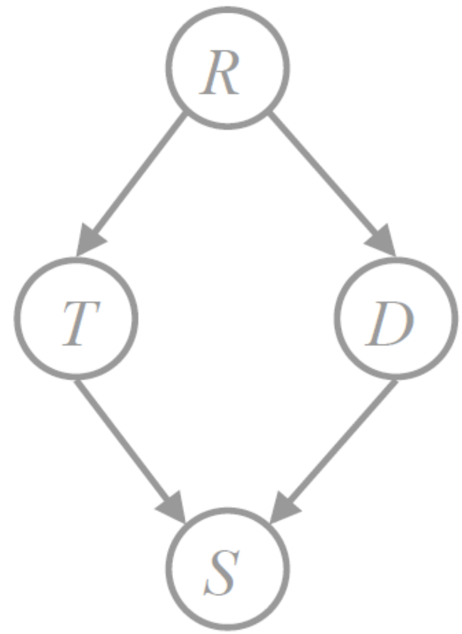


Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I' m sad
- Questions:

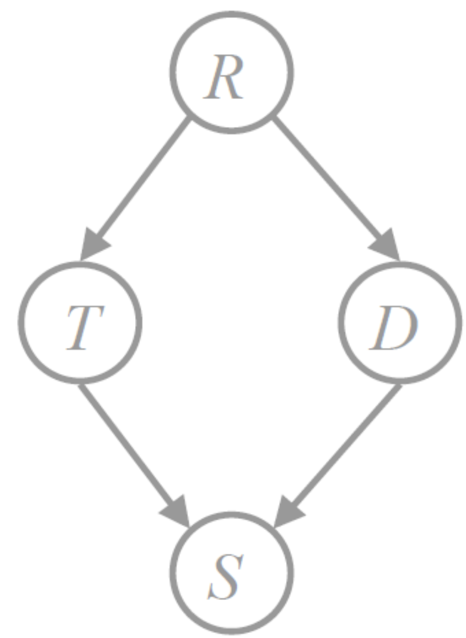
$T \perp\!\!\!\perp D$

$T \perp\!\!\!\perp D | R$ Yes



Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I' m sad
- Questions:



$$T \perp\!\!\!\perp D$$

$$T \perp\!\!\!\perp D | R$$ Yes

$$T \perp\!\!\!\perp D | R, S$$

All Conditional Independences

- Given a Bayes net structure, can run d-separation to build a complete list of conditional independences that are guaranteed to be true, all of the form

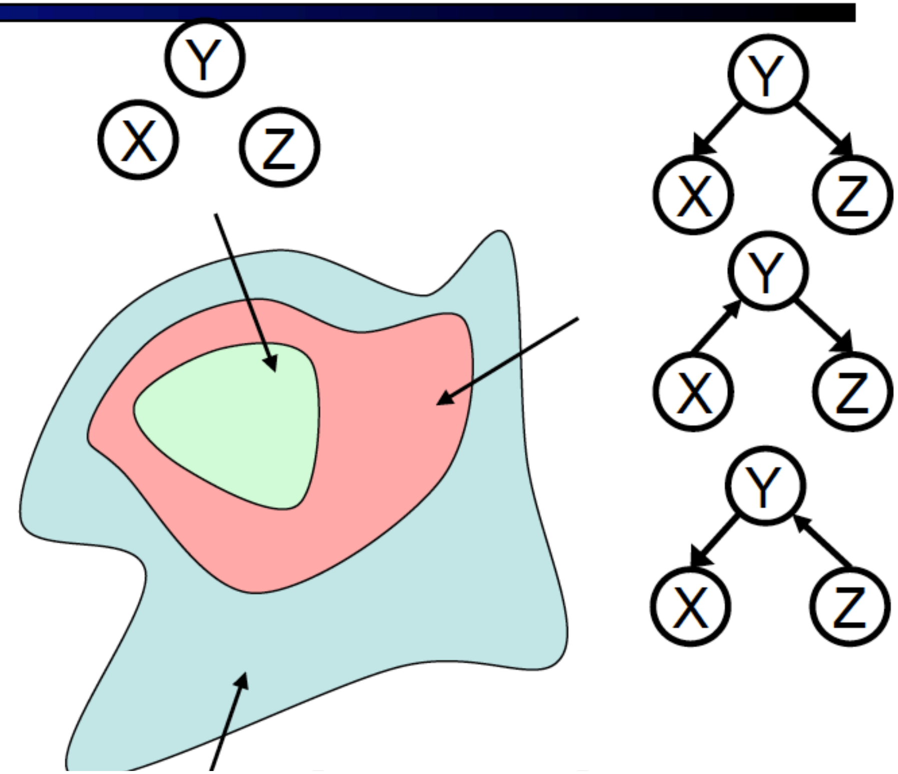
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded

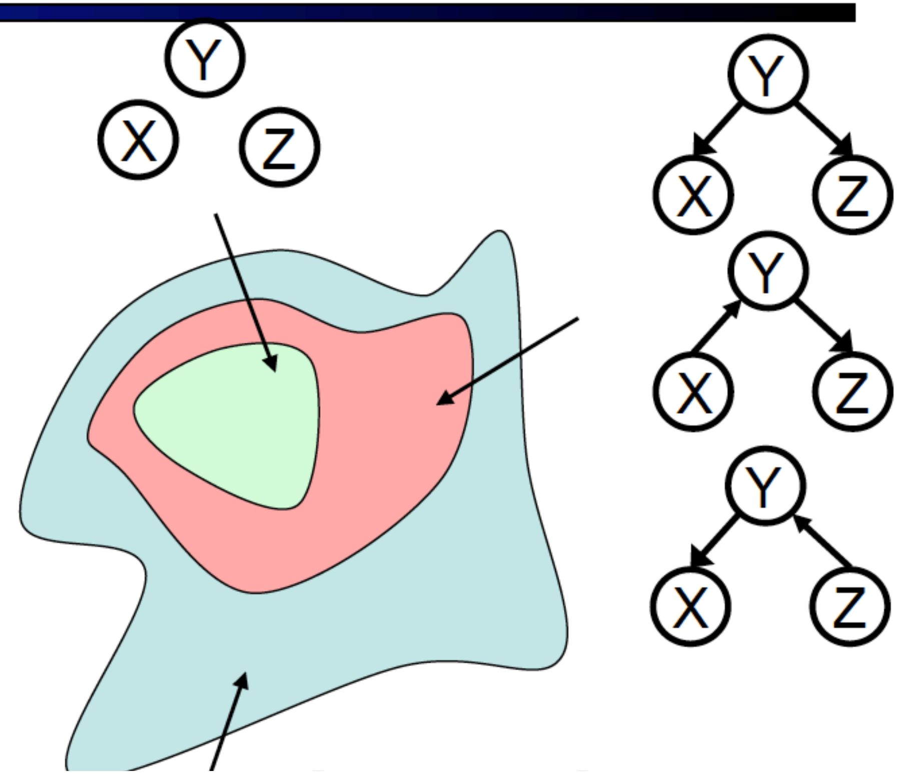
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences



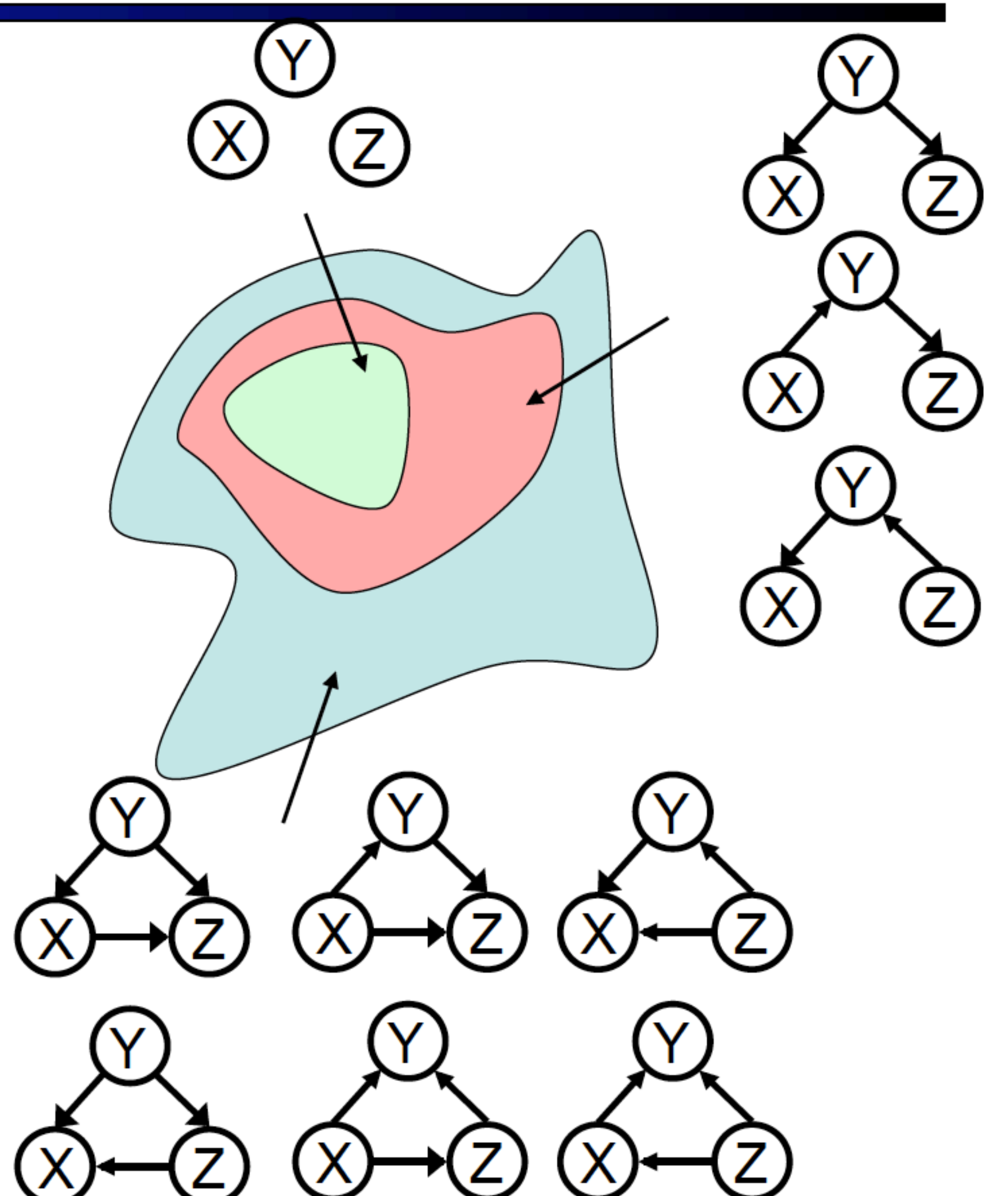
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)



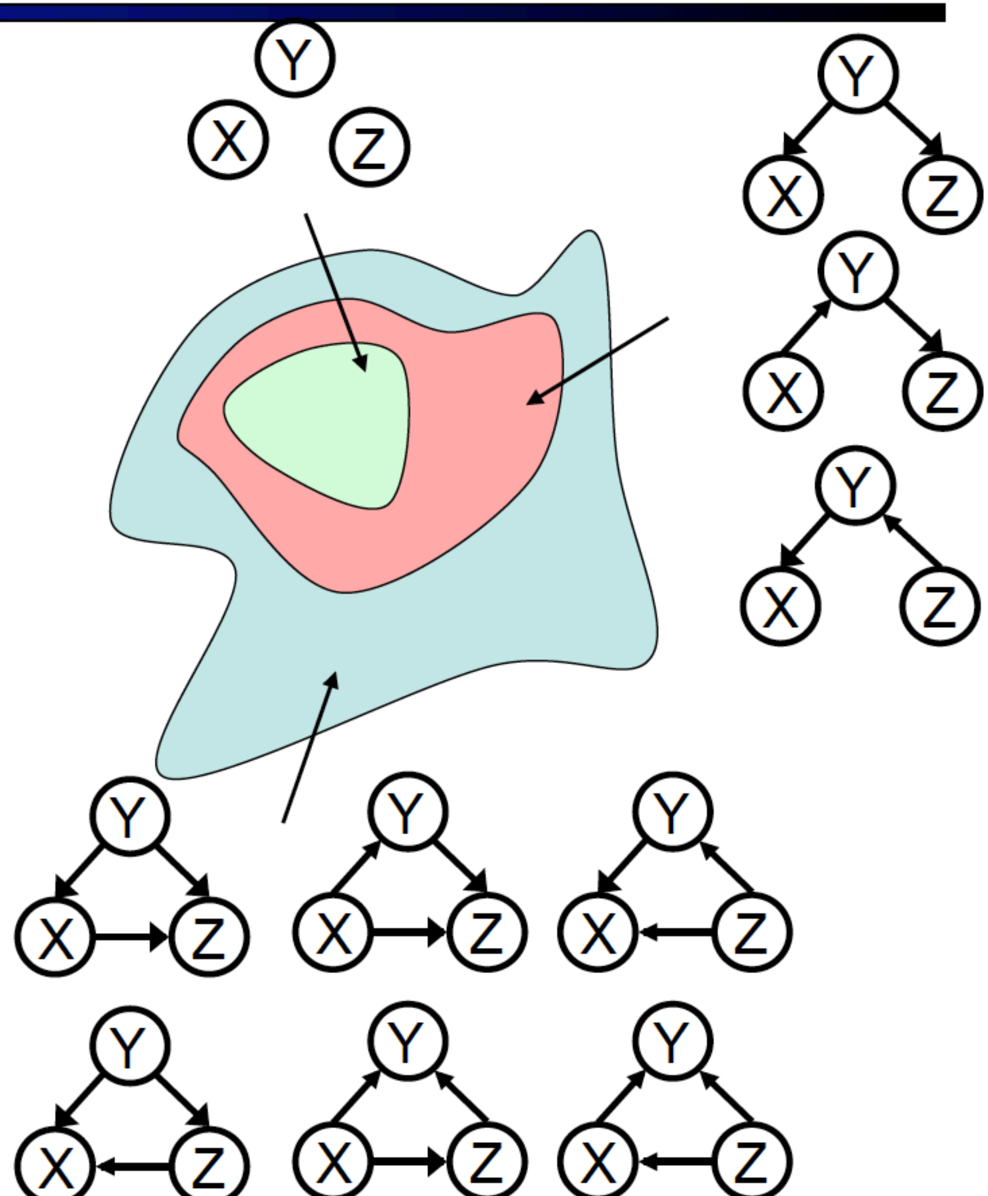
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs



Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions

- Guaranteed independencies of distributions can be deduced from BN graph structure

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

Bayes' Nets

- ✓ Representation
- ✓ Conditional Independences
- Probabilistic Inference
  - Enumeration (exact, exponential complexity)
  - Variable elimination (exact, worst-case exponential complexity, often better)
  - Probabilistic inference is NP-complete
  - Sampling (approximate)
- Learning Bayes' Nets from Data

Possible to have same full list of conditional independencies for different BN graphs?

- Yes!
- Examples:

- If two Bayes' Net graphs have the same full list of conditional independencies then they are able to encode the same set of distributions.