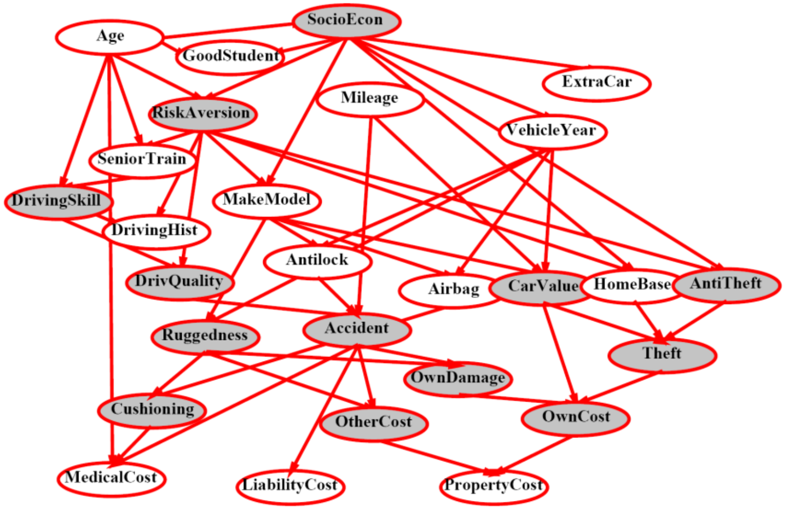
贝叶斯网络: 独立性(INDEPENDENCE)

概率复习

- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Chain rule $$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \qquad X \perp\!\!\!\perp Y | Z$$

Bayes' Nets 贝叶斯网络

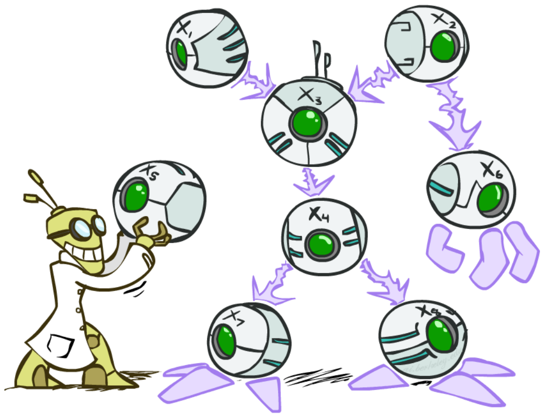


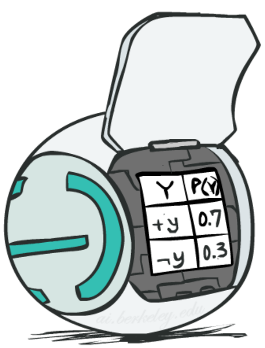
- 对一个领域里变量构成的概率模型的有效率的编码
- Questions we can ask:
  - 推理: given a fixed BN, what is P(X | e)?
  - 表达: given a BN graph, what kinds of distributions can it encode?
  - 建模: what BN is most appropriate for a given domain?

贝叶斯网络的语义

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values $P(X|a_1 \ldots a_n)$
- 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:

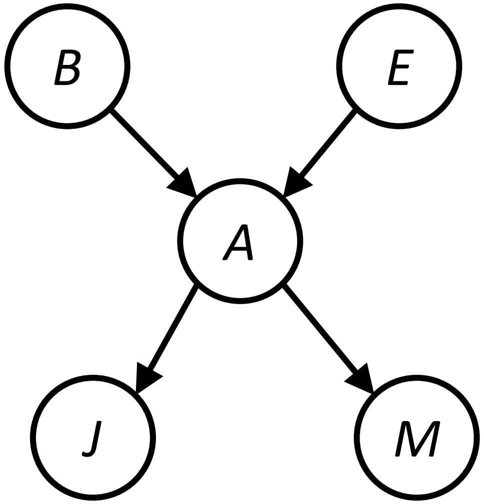
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$



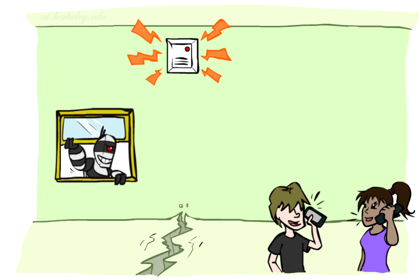


Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



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

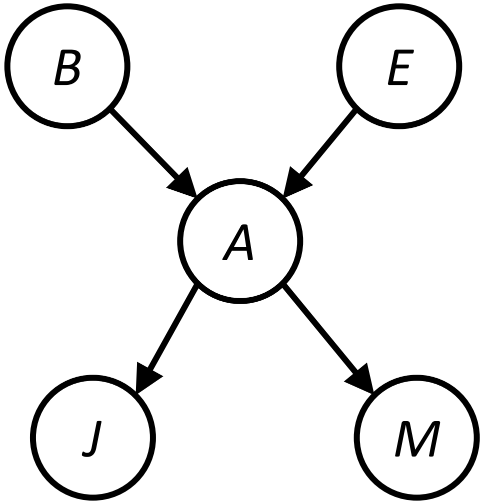
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

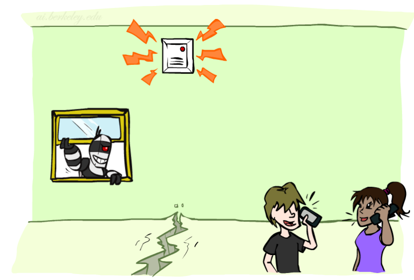
$$P(+b, -e, +a, -j, +m) =$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



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

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

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Size of a Bayes' Net

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

  2^N

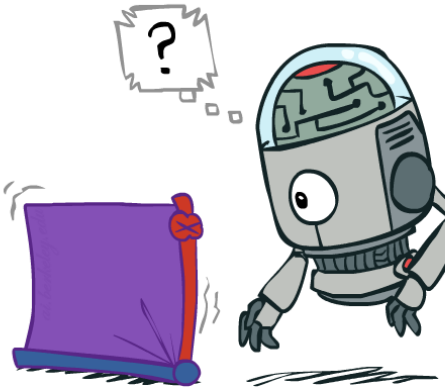
- How big is an N-node net if nodes have up to k parents?

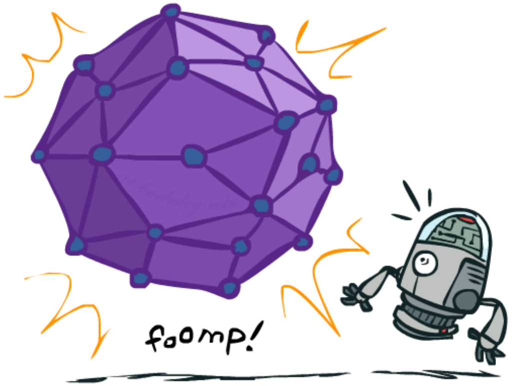
  $O(N * 2^{k+1})$

- 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 faster to answer queries (coming)





Bayes' Nets

- ✓ Representation
- Conditional Independences
- Probabilistic Inference
- Learning Bayes' Nets from Data

Conditional Independence 条件独立性

- X and Y are independent if

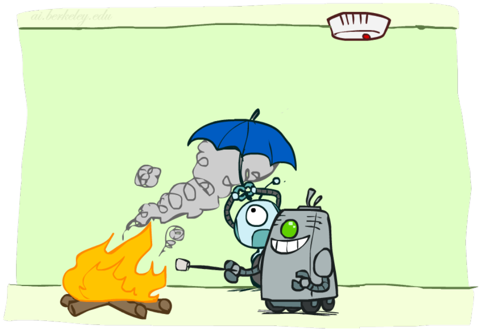
$$\forall x, y \;\; P(x, y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

- X and Y are conditionally independent given Z

$$\forall x, y, z \;\; P(x, y|z) = P(x|z)P(y|z) \dashrightarrow X \perp\!\!\!\perp Y|Z$$

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp\!\!\!\perp Fire|Smoke$

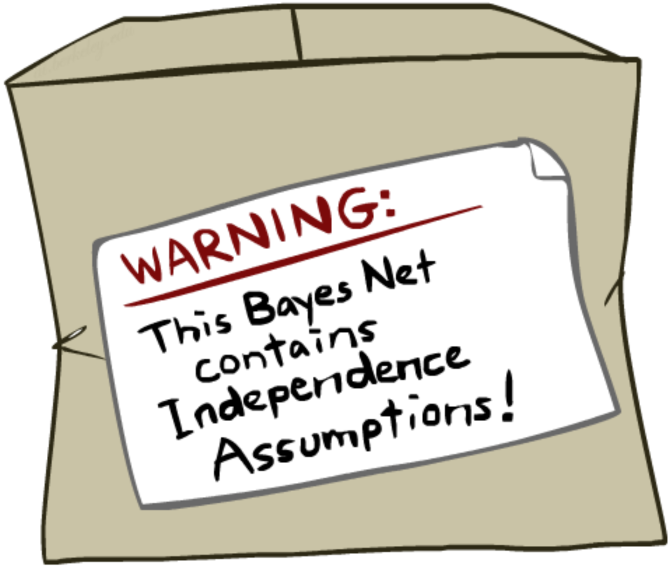


Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:

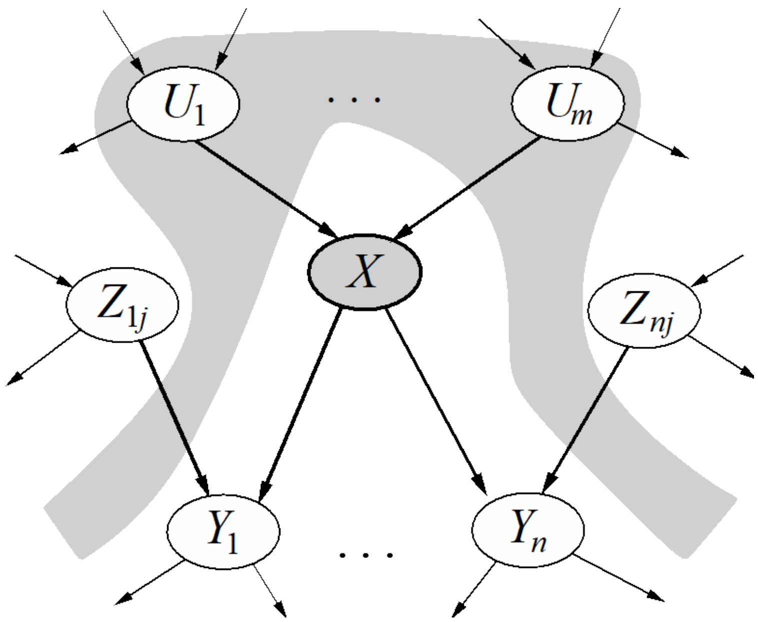
$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Beyond above "chain rule → Bayes net" conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph



条件独立性语义

- *当给定它的父节点取值后，每个变量都是条件独立于它的其他祖先节点变量*



贝叶斯网络里的概率

- 为什么我们可以保证以下公式是正确的联合分布

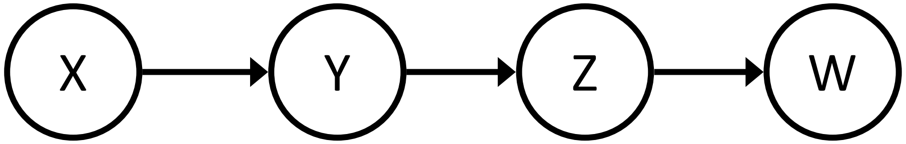
$$P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$$

- 连锁法 (对所有分布有效): $P(X_1,..,X_n) = \prod_i P(X_i \mid X_1,...,X_{i-1})$
- <u>假定</u> 条件独立性: $P(X_i \mid X_1,...,X_{i-1}) = P(X_i \mid Parents(X_i))$
  - 当加入节点 X_i, 确保其父节点 “屏蔽” 它与其他祖先节点的联系
  - 给定它的父节点，每个变量条件独立于它的非子孙节点变量(即所有它之前的变量)

→ 结果: $P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$

- 所以，网络的拓扑结构暗示着条件独立性

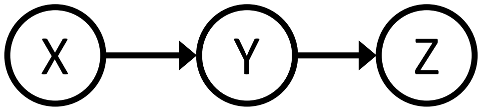
Example



- 蕴含哪些条件独立性假设 （directly from simplifications in chain rule ）？
- 还有没有蕴含其他的条件独立性假设?

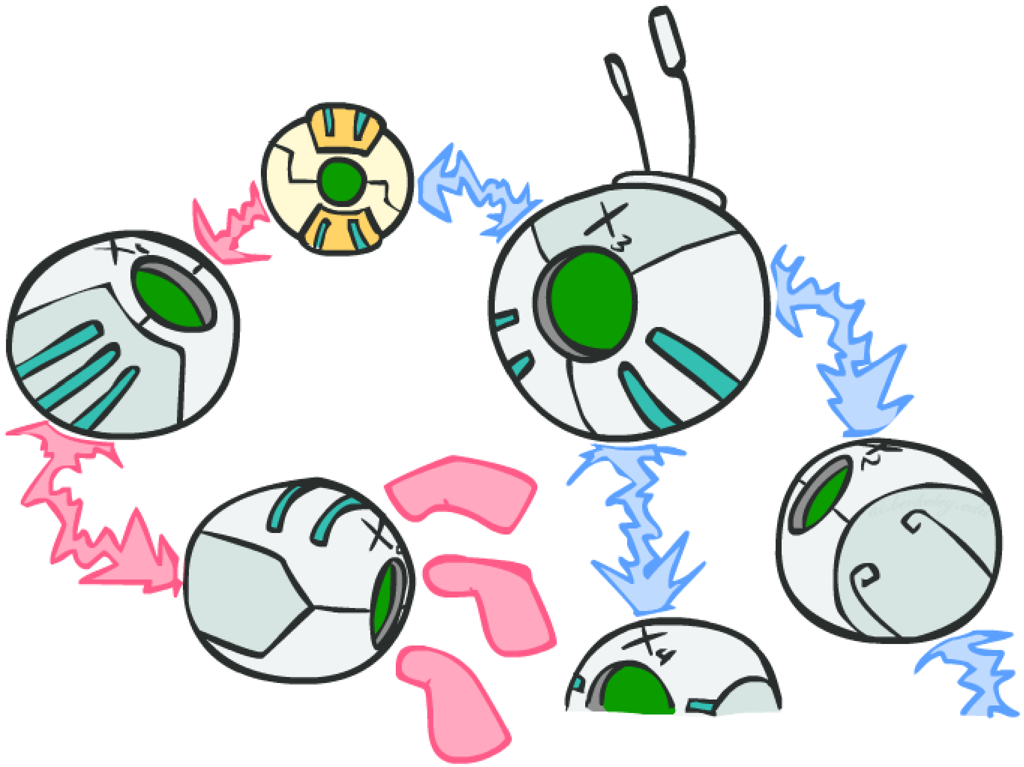
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:



  - Question: are X and Z necessarily independent?
    - Answer: no. Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Addendum: they *could* be independent: how?

D-separation(D分离): Outline



Causal Chains（因果关系链）

- This configuration is a “causal chain”

X: Low pressure　　Y: Rain　　Z: Traffic

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

- Guaranteed X 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:
    - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
    - In numbers:

      P(+y | +x) = 1, P(-y | - x) = 1,
      P(+z | +y) = 1, P(-z | -y) = 1

Causal Chains（因果关系链）

- This configuration is a “causal chain”

X: Low pressure   Y: Rain   Z: Traffic

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

- Guaranteed X independent of Z given Y?

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

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

$$= P(z|y)$$

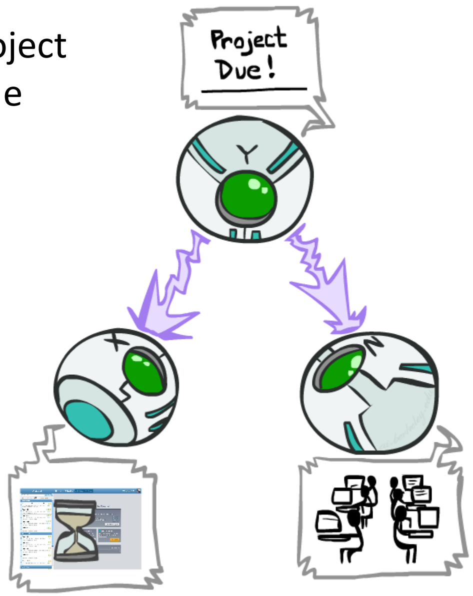
Yes!

- Evidence along the chain “blocks” the influence

Common Cause（原因相同）

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

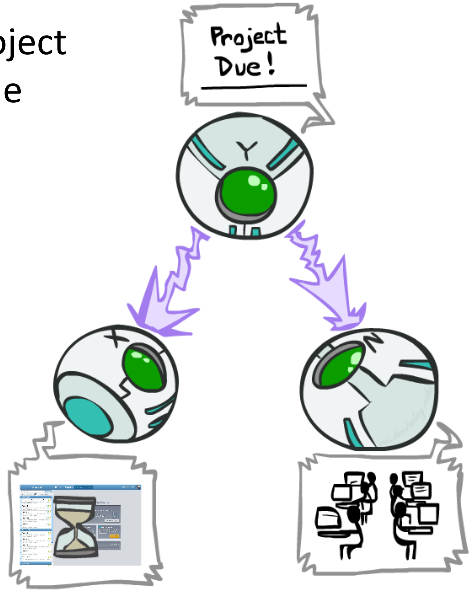
$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X 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:
    - Project due causes both forums busy and lab full
    - In numbers:

      P(+x | +y) = 1, P(-x | -y) = 1,
      P(+z | +y) = 1, P(-z | -y) = 1

Common Cause（原因相同）

- This configuration is a "common cause"



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

- Guaranteed X and Z independent given Y?

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

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

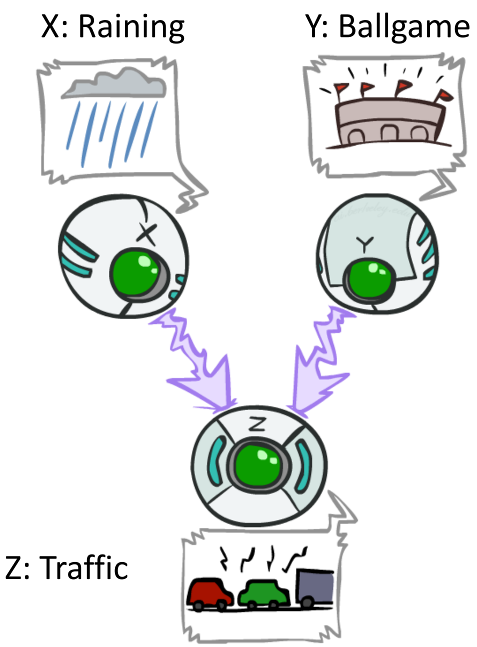
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect（结果相同）

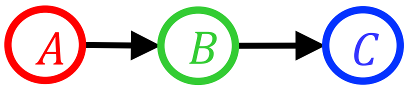
- Last configuration: two causes of one effect (v-structures)



- Are X and Y 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 Y independent given Z?
  - *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.

条件独立语法

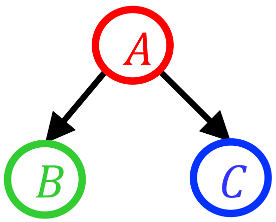
- 对于下列贝叶斯网络, 写出联合分布 $P(A,B,C)$
  1. 使用链式法则　(顺序为A,B,C)
  2. 使用贝叶斯网络语法 (CPT 的乘积)



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

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

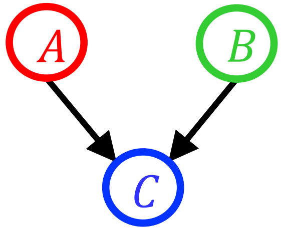
前提:
$P(C|A,B) = P(C|B)$
C 独立于A 当给定B后



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

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

前提:
$P(C|A,B) = P(C|A)$
B 和 C 独立 给定A的值后



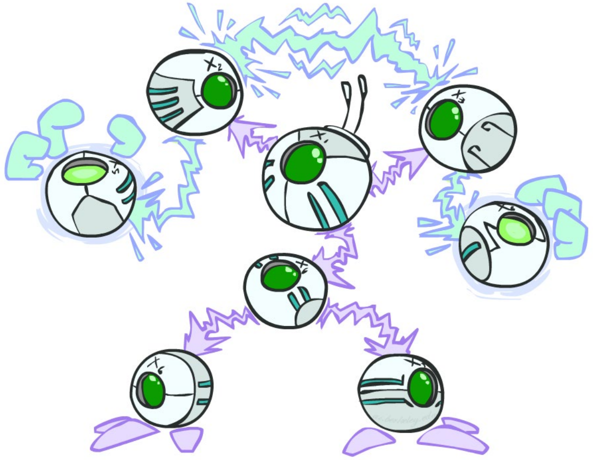
$P(A)\ P(B|A)\ P(C|A,B)$

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

前提:
$P(B|A) = P(B)$
A和B独立

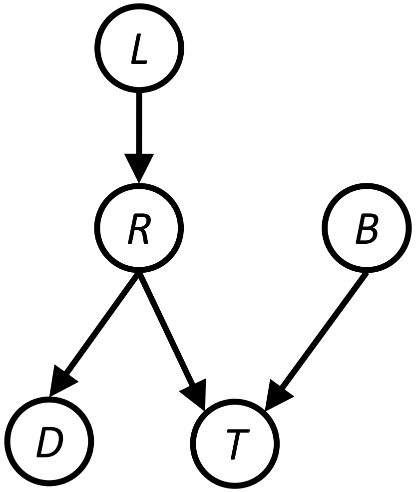
The General Case(一般情况)

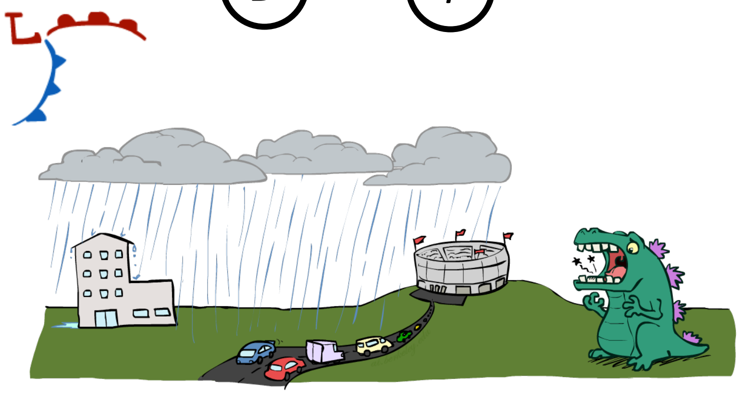
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph
- Any complex example can be broken
  into repetitions of the three canonical cases



Reachability（联通性判断是否条件独立）

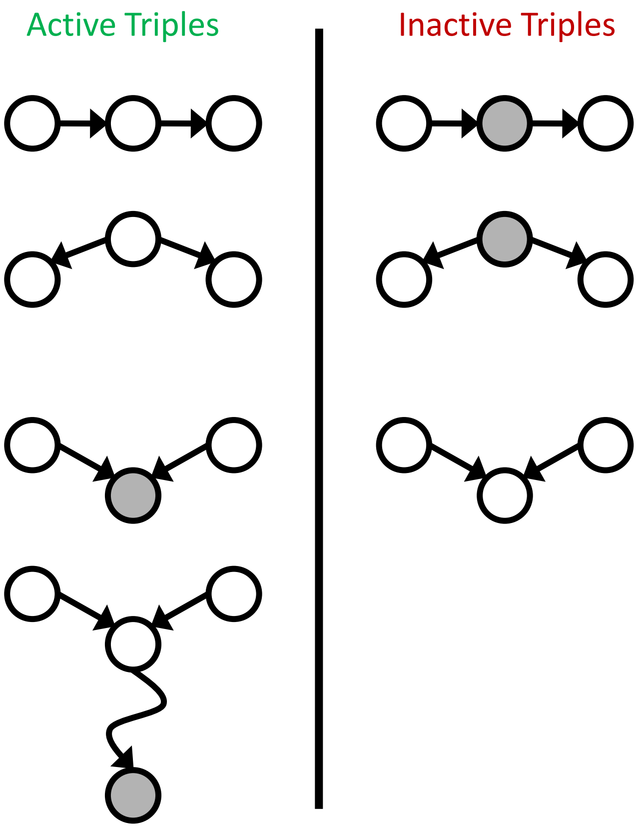
- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't count as a link in a path unless "active"





Active / Inactive Paths

- Question: Are X and Y conditionally independent given evidence variables {Z}?
  - Yes, if X and Y "d-separated" by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is 通路 if each triple is active:
  - Causal chain A → B → C where B is unobserved (either direction)
  - Common cause A ← B → C where B is unobserved
  - Common effect (aka v-structure)
    A → B ← C where B *or one of its descendents* is observed
- All it takes to block a path is a single inactive segment



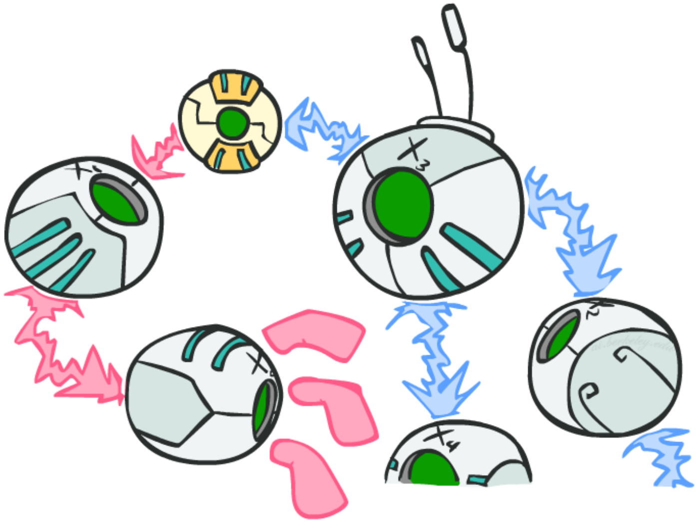
D-Separation（D分离）

- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$?
- Check all (undirected!) paths between X_i and X_j
  - If one or more active, then independence not guaranteed

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

  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

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

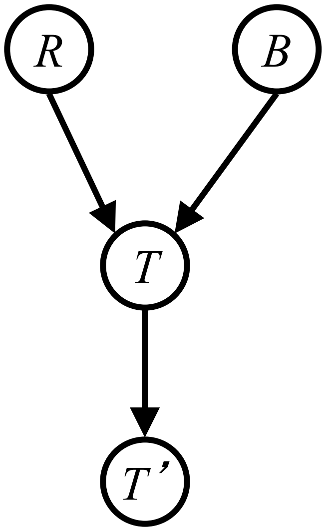


Example

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

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

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



Example

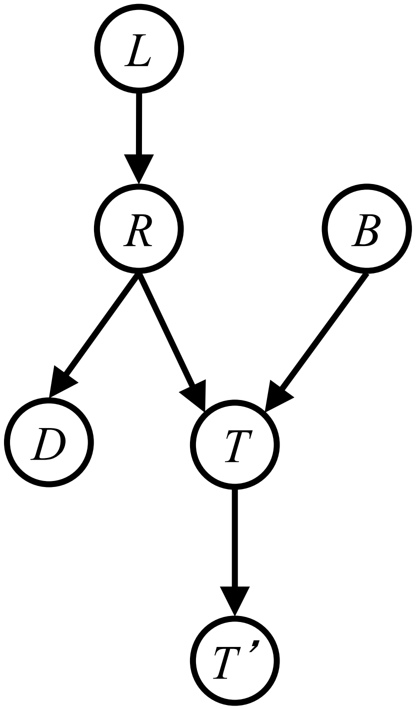
$L \perp\!\!\!\perp T' | T$ *Yes*

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

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

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

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



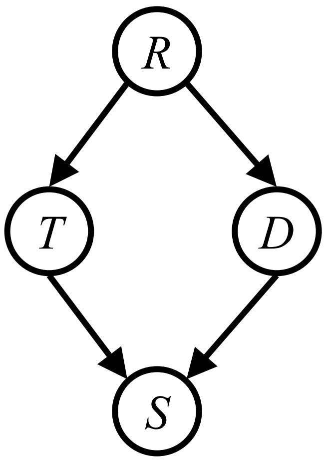
Example

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

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

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

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

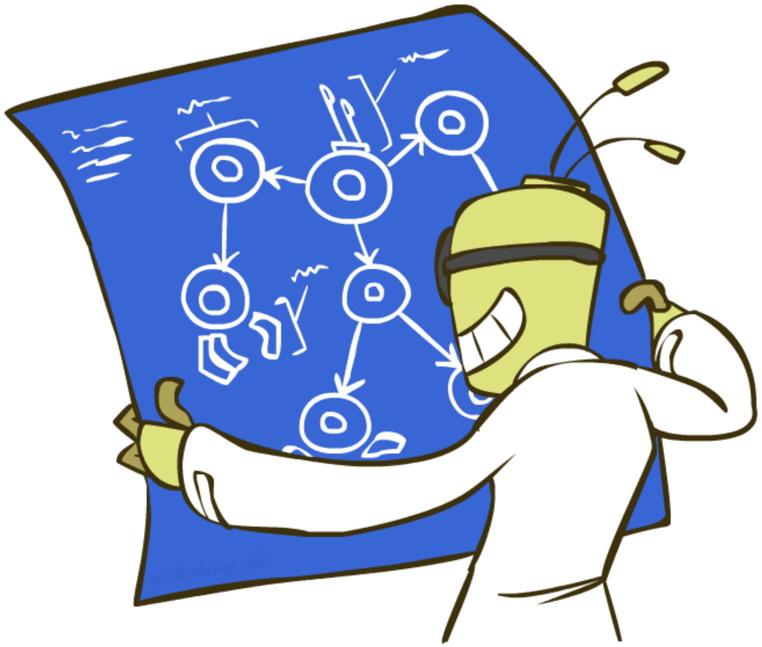


图形结构蕴含了条件独立假设

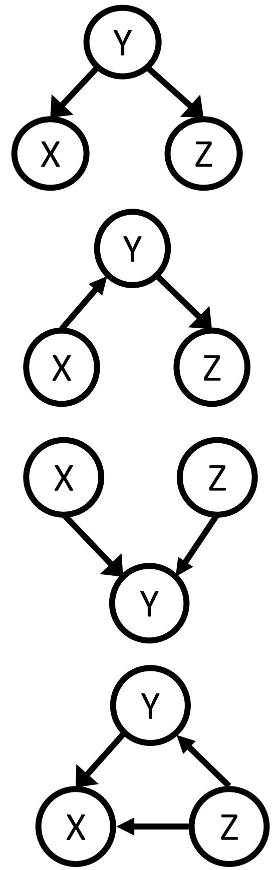
- 给定一个贝叶斯网络结构，可以运行 d-separation 算法，找到所有条件独立假设：

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

- This list determines the set of probability distributions that can be represented

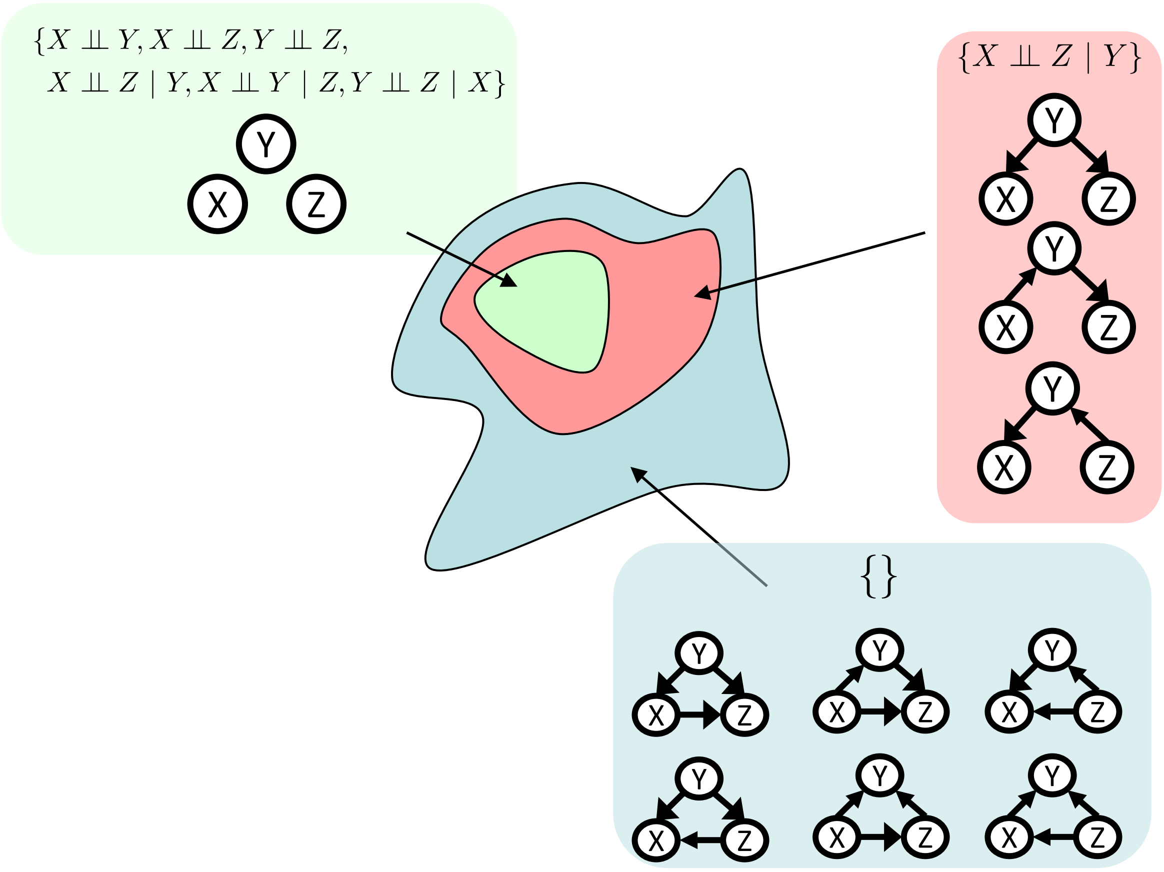


尝试计算所有表达的独立性假设



图结构限制了所能表达的分布集合

- Given some graph topology G, only certain joint distributions can be encoded
- 图结构保证了（条件）独立假设的存在
- (There might be more independence)
- 增加边，扩大了所能表示的分布集合，but has several costs
- Full conditioning can encode any distribution



贝叶斯网络表达语义小结

- 贝叶斯网络 compactly encode 联合分布
- 从贝叶斯网络图的结构可以推出独立性假设
- D-separation gives precise 条件独立性假设 from graph alone
- 一个贝叶斯网络所表达的联合分布模型中可能还存在（D分离）监测不出来的（条件）独立情况，除非这时你检查具体的分布中的数值（来做进一步的推断）

贝叶斯网络内容

- ✓ 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