COMS W4701: Artificial Intelligence

Lecture 19: Bayesian Networks

Tony Dear, Ph.D.
Department of Computer Science
School of Engineering and Applied Sciences

Today

- Bayesian networks
- Bayes net semantics
- Conditional independences
- D-separation

Probabilistic Graphical Models

- Probabilistic models can encode knowledge and associated uncertainty, including exceptions and special cases without full enumeration
- System aspects are captured by joint distributions over random variables

- A graphical model uses graphs to compactly encode a complex distribution
- It also represents *factorizations* that can be used to simplify the model

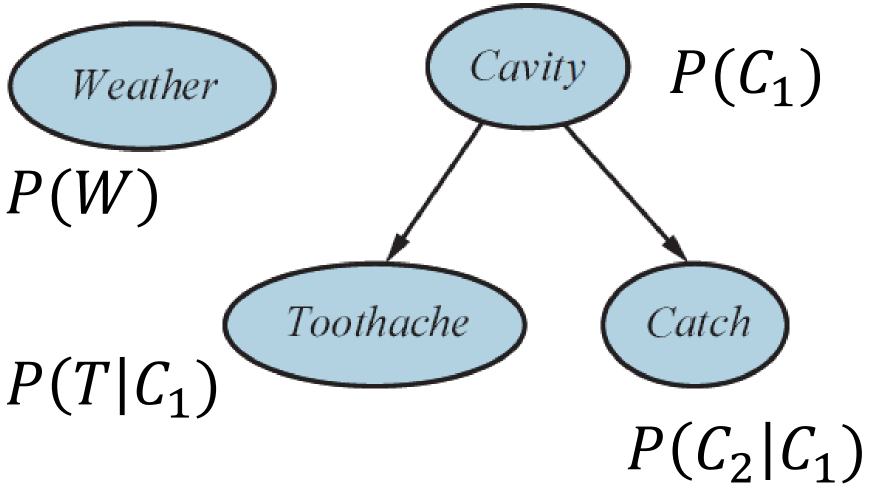
- Such models are more easily interpretable and transparent for users
- Are more amenable to *inference* and *learning* for model construction

Bayesian Networks

- A **Bayesian network** is a directed acyclic graph (DAG) representing a joint distribution
- Captures both a factorization as well as a set of conditional independences

- Each node corresponds to a random variable

- Each edge indicates influence or correlation
- May also be causation, but not always



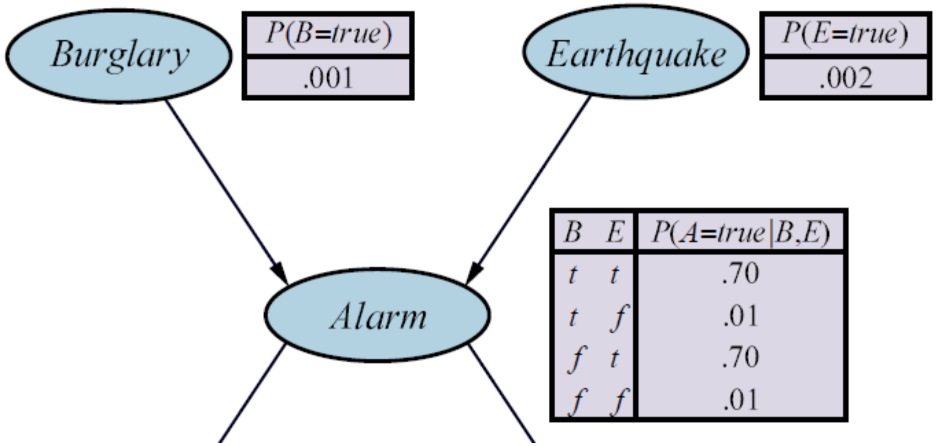
- **Parameters** of the Bayes net: A *local* conditional probability table (CPT) for each node
- The CPT for node X_i contains the values $P(X_i|parents(X_i))$

Conditional Probability Tables

- A CPT contains *all* possible conditional distributions $P(X_i|parents(X_i))$

- If each RV domain is size d and X_i has k parents, then there are d^k combinations of parent values, d^k different conditional distributions
- If X_i is also size k, then CPT has d^{k+1} parameters in total

- Optimization for CPTs of binary RVs:
- Can simply store half of the parameters



Joint Distribution

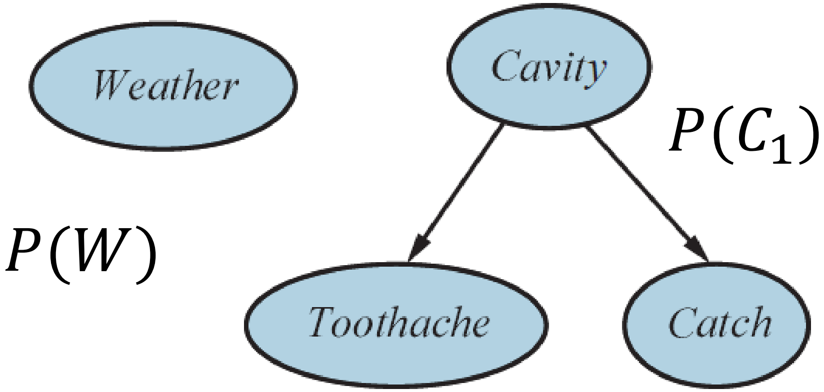
- *Assumption*: X_i is conditionally independent of its non-descendants given its parents

- Given a **topological ordering** of nodes $X_1, \dots, X_n$ s.t. all ancestors of a node occur before it, Bayes net joint probabilities are defined as follows:

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

- Example calculations:

$$P(w, c_1, t, c_2) = P(w)P(c_1)P(t|c_1)P(c_2|c_1)$$
$$= P(c_1)P(c_2|c_1)P(t|c_1)P(w)$$



$P(C_1)$

$P(W)$

$P(T|C_1)$

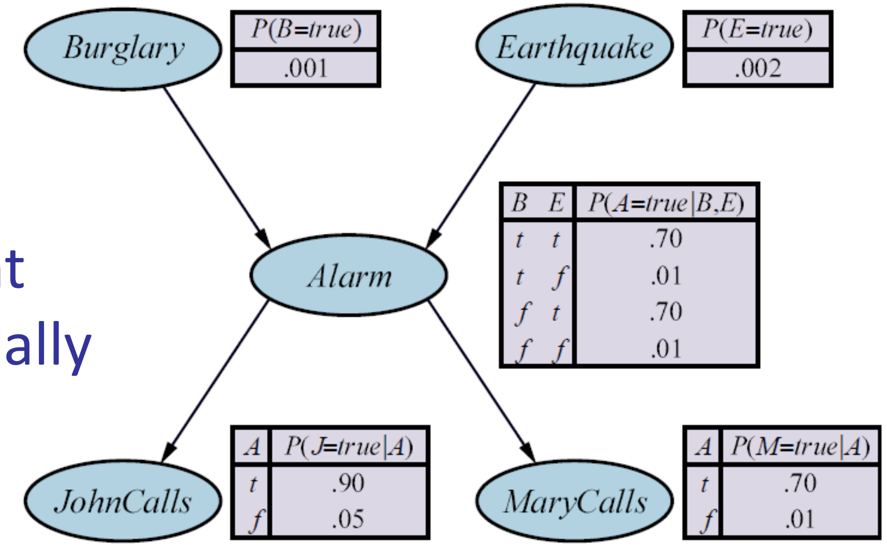
$P(C_2|C_1)$

Constructing Bayes Nets

- Given a set of random variables, *choose* a topological ordering for them
- Add a new node to the network for each RV following this ordering

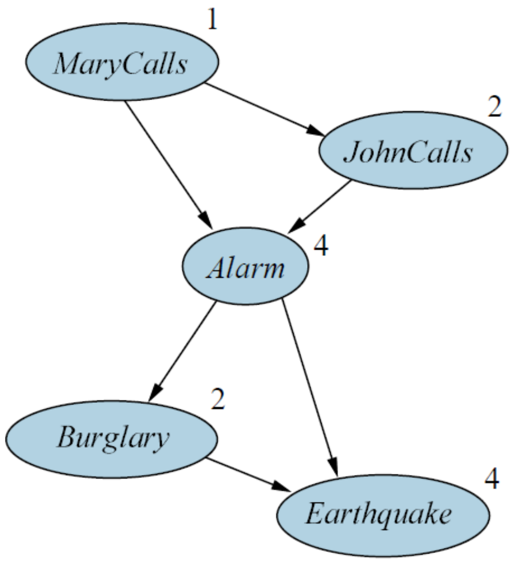
- When adding X_i, each previous node $X_j, j \in \{1, \ldots, i-1\}$, is a direct *parent* if X_i and X_j are *not* conditionally independent given other parents

- Ex: *Burglary* and *Earthquake* conditionally independent given nothing, *JohnCalls* and *MaryCalls* each conditionally independent of all other RVs given *Alarm*

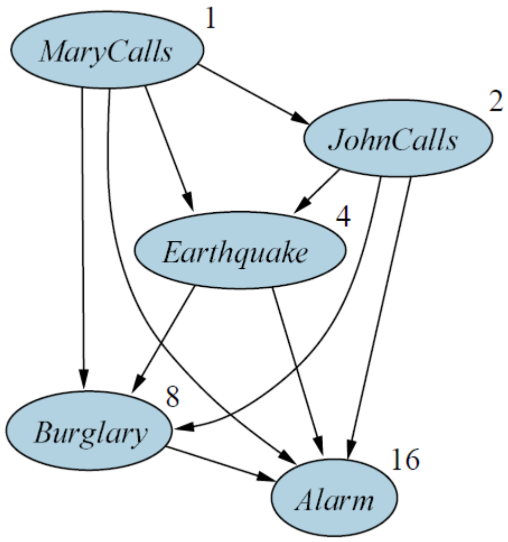


Example: Alarm Network

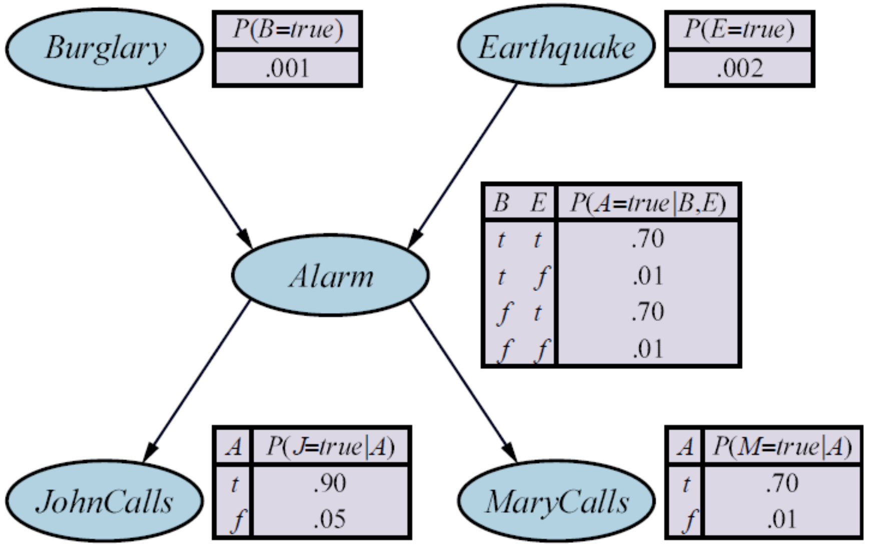
- The chosen ordering can drastically affect the size of the network!
- Prefer topological orderings (e.g., cause -> effect) that result in nodes with fewer parents / dependencies, or more *compact* networks



13 parameters

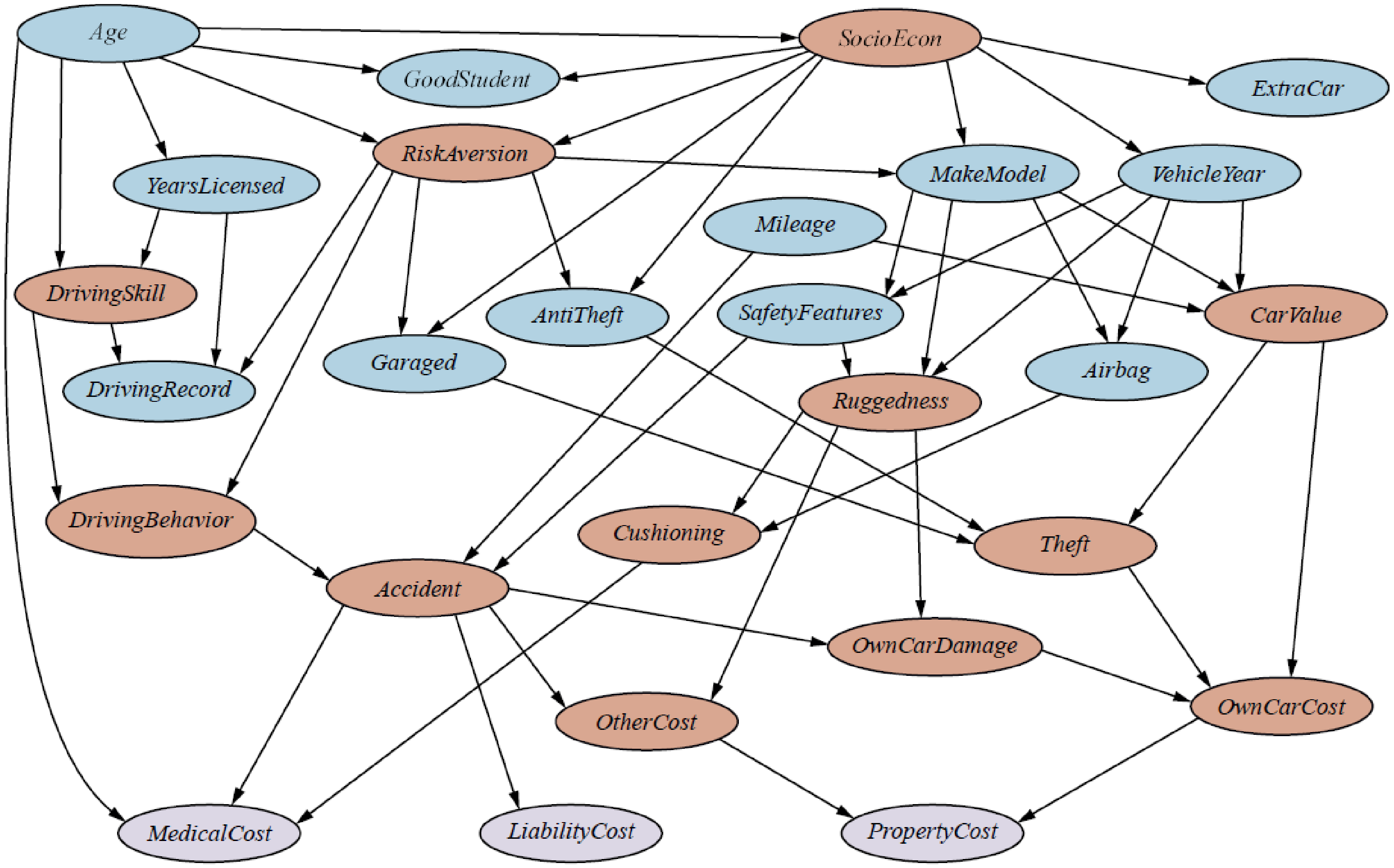


31 parameters



10 parameters

Example: Car Insurance

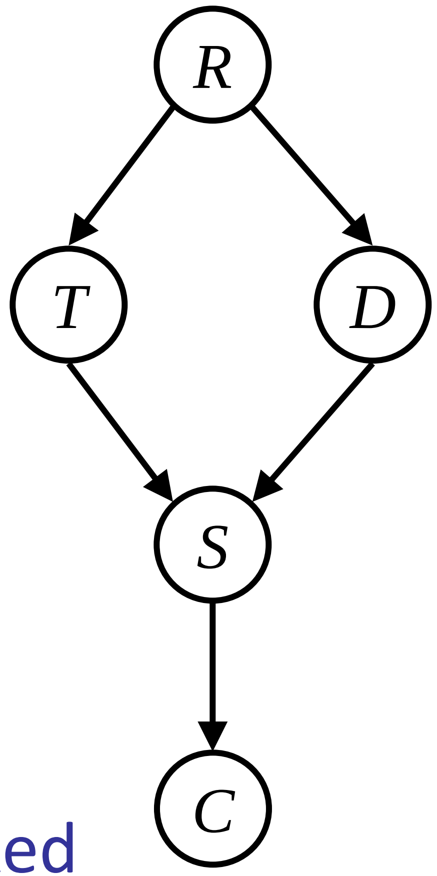


Inferring Conditional Independence

- Recall: A node X is conditionally independent of all non-descendants given observed values of all its parents
- Think of observed nodes as *blocking information flow*

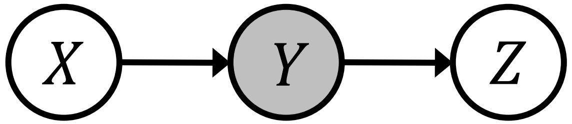
- We can *extend* this independence guarantee to other pairs of nodes if observed nodes also block all paths between them
- Examine local structures of 3 nodes (2 edges) at a time

- X_i and X_j are independent if all paths between them are blocked



Chains and Forks

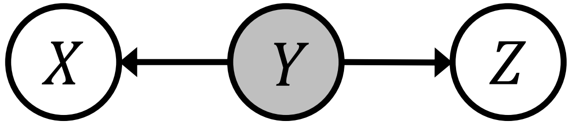
- Generally, nodes X and Z in **chain** and **fork** structures are not independent



$P(X)$ $P(Y|X)$ $P(Z|Y)$

- If Y is observed, then path between X and Z is blocked and they *become* conditionally independent

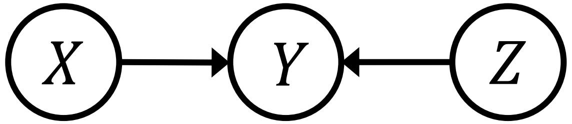
- If removing Y breaks the network into two components, all nodes in X's component become conditionally independent of all nodes in Z's



$P(X|Y)$ $P(Y)$ $P(Z|Y)$

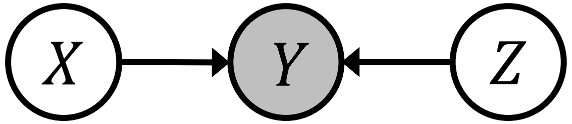
Colliders

- If X and Z share only **colliders** (descendants), the pair is guaranteed to be independent if no colliders are observed



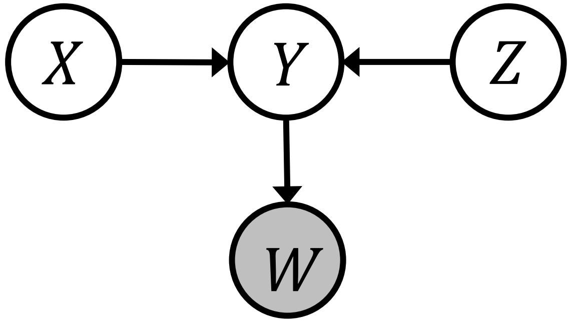
$P(X)$ $\quad P(Y|X,Z)$ $\quad P(Z)$

- But X and Z are *not* guaranteed conditionally independent given observation of a collider!



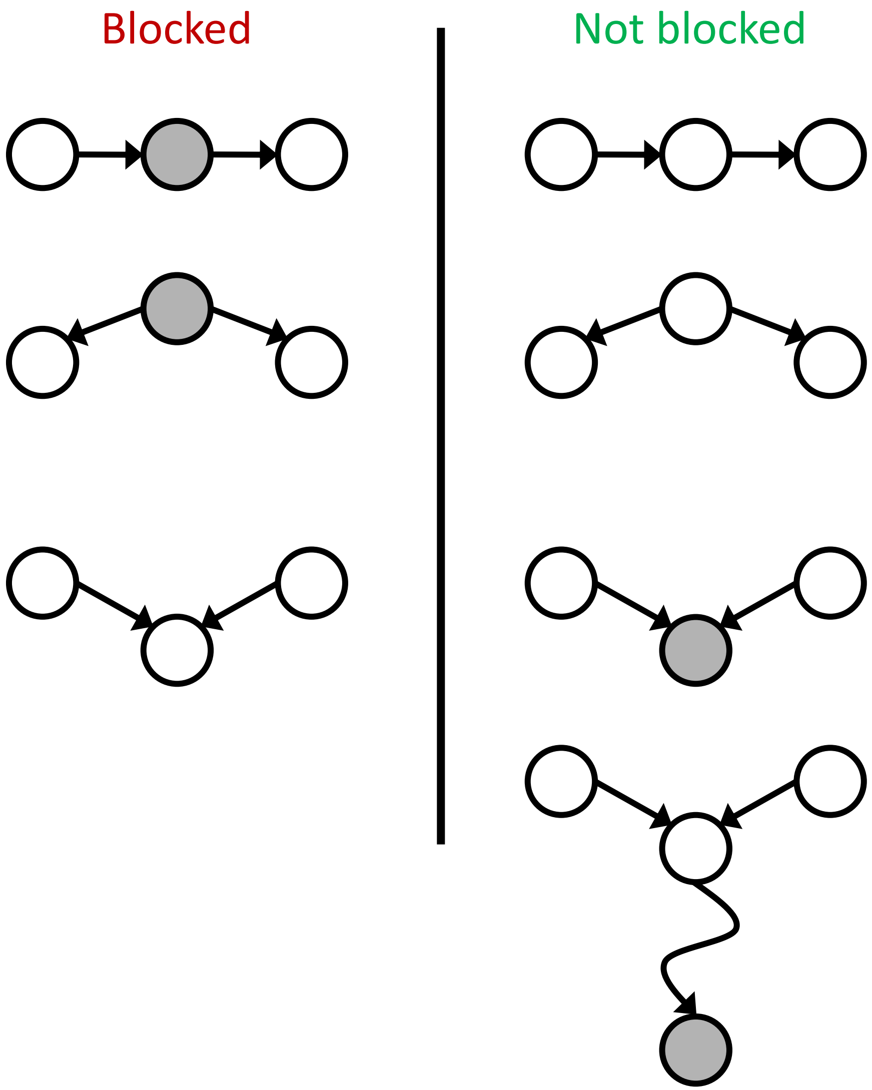
- $P(x,z|y) = \frac{P(x,y,z)}{P(y)} = \frac{P(x)P(z)P(y|x,z)}{P(y)}$
- $P(x|y)P(z|y) = \frac{P(y|x)P(x)}{P(y)}\frac{P(y|z)P(z)}{P(y)}$

Generally not equal!



D-Separation

- To check whether X_i and X_j are conditionally independent given a set of observed nodes Z:
- Check every possible path between X_i and X_j in the "undirected" version of the Bayes net
- **Independent** and "d-separated" if *every* path is blocked; otherwise, **not guaranteed independent** if *at least one* path is not blocked



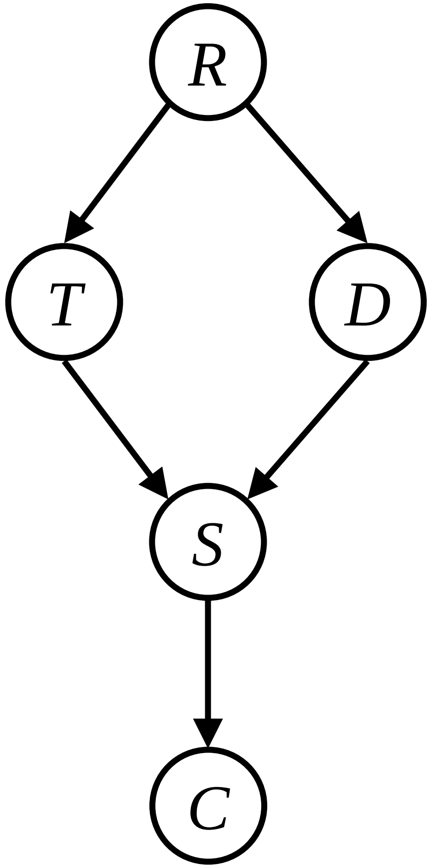
Example: D-Separation

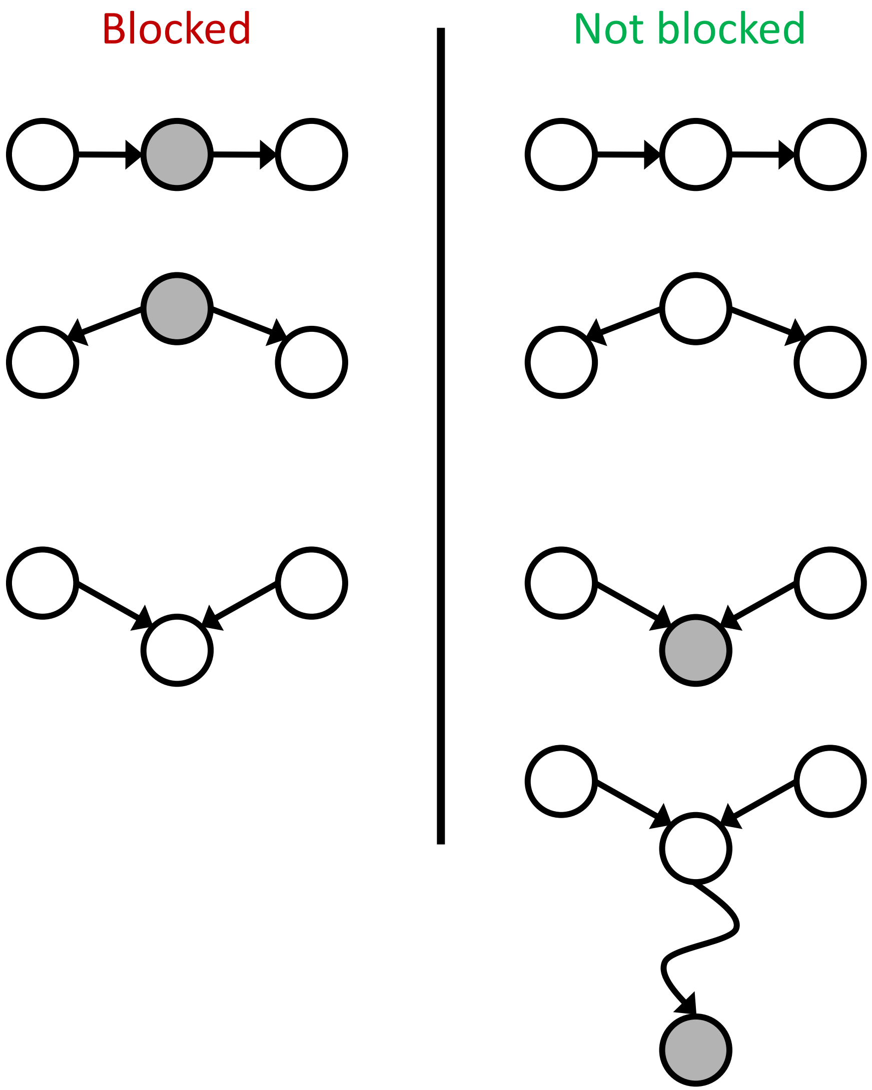
Which nodes are independent…

- Given S?
- Given R?

- Given T or D?
- Given T and D?

- Given R and S?
- Given R and C?





Summary

- Bayesian networks graphically encode independence assumptions about joint distributions in a compact way

- Bayes net representations of a joint distribution are not unique, but we can design them to be compact and minimize the number of parameters

- D-separation rules can help infer local independences given evidence
- Consider whether information can “flow” from one node to another