CMSC 471:
Reasoning with Bayesian Belief Network

Chapters 12 & 13

KMA Solaiman – ksolaima@umbc.edu

Some slides courtesy Tim Finin and Frank Ferraro

Overview

- Bayesian Belief Networks (BBNs) can reason with networks of propositions and associated probabilities
- Useful for many AI problems
  - Diagnosis
  - Expert systems
  - Planning
  - Learning

Probabilistic Graphical Models

A graph G that represents a probability distribution over N random variables $X_1, \dots, X_N$

Probabilistic Graphical Models

A graph G that represents a probability distribution over N random variables $X_1, \ldots, X_N$

Graph G = (vertices V, edges E)

Distribution $P(X_1, \ldots, X_N)$

Probabilistic Graphical Models

A graph G that represents a probability distribution over random N variables $X_1, \dots, X_N$

Graph G = (vertices V, edges E)

Distribution $P(X_1, \dots, X_N)$

Vertices ↔ random variables

Edges show dependencies among random variables

Probabilistic Graphical Models

A graph G that represents a probability distribution over N random variables $X_1, \dots, X_N$

Graph G = (vertices V, edges E)
Distribution $p(X_1, \dots, X_N)$

Vertices ↔ random variables
Edges show dependencies among random variables

Two main flavors: *directed* graphical models and *undirected* graphical models

Probabilistic Graphical Models

A graph G that represents a probability distribution over N random variables $X_1, \ldots, X_N$

Graph G = (vertices V, edges E)
Distribution $p(X_1, \ldots, X_N)$

Vertices ↔ random variables
Edges show dependencies among random variables

Two main flavors: ***directed*** **graphical models** and *undirected* graphical models

Directed Graphical Models

A *directed* (acyclic) graph G=(V,E) that represents a probability distribution over random variables $X_1, \dots, X_N$

Joint probability factorizes into factors of X_i conditioned on the parents of X_i

Directed Graphical Models

A *directed* (acyclic) graph G=(V,E) that represents a probability distribution over random variables $X_1, \dots, X_N$

Joint probability factorizes into factors of X_i conditioned on the parents of X_i

Benefit: the independence properties are *transparent*

Directed Graphical Models

A *directed* (acyclic) graph G=(V,E) that represents a probability distribution over random variables $X_1, \ldots, X_N$

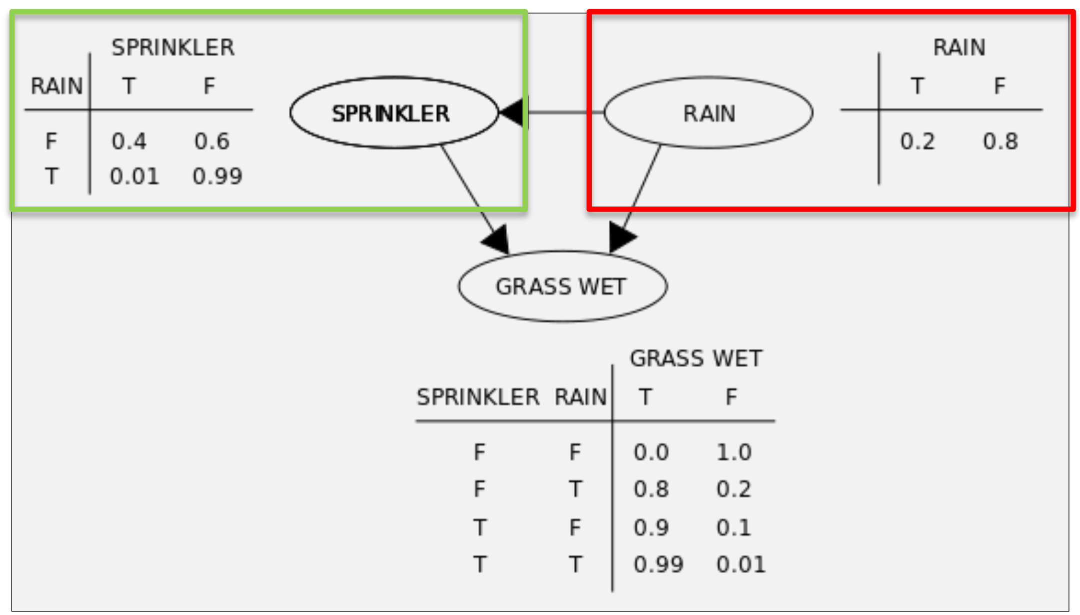
Joint probability factorizes into factors of X_i conditioned on the parents of X_i

A graph/joint distribution that follows this is a **Bayesian network**

BBN Definition

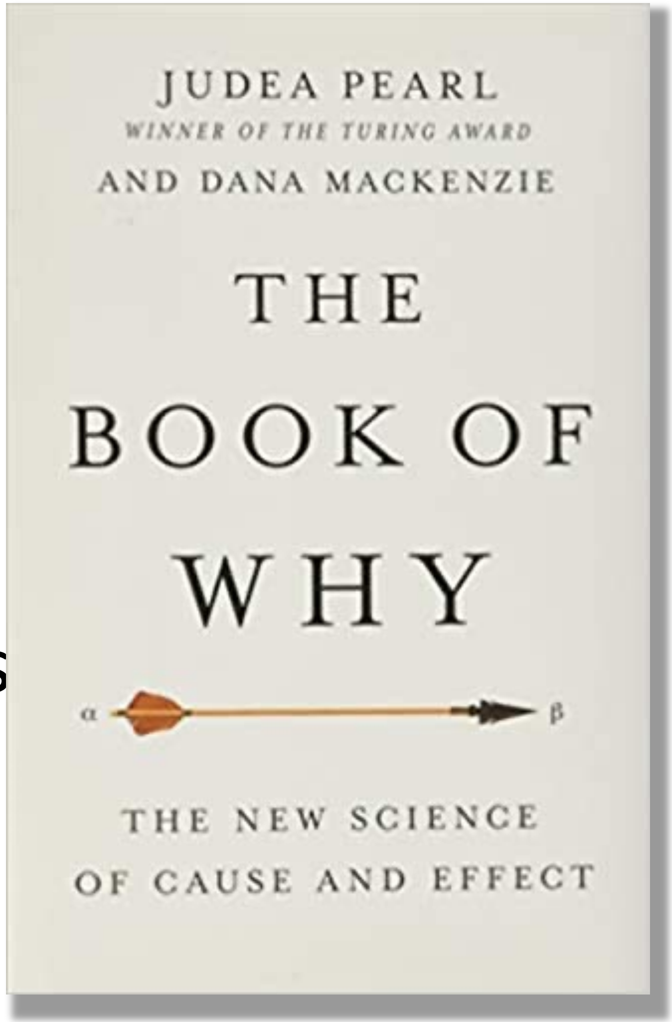
- AKA Bayesian Network, Bayes Net
- A graphical model (as a [DAG]) of probabilistic relationships among a set of random variables
- Nodes are variables, links represent direct influence of one variable on another
- Nodes have **prior probabilities** or **conditional probability tables** (CPTs)

[source]



History lesson: Judea Pearl

- UCLA CS professor
- Introduced Bayesian networks in the 1980s
- Pioneer of probabilistic approach to AI reasoning
- First to formalize causal modeling in empirical sciences
- Written many books on the topics, including the popular 2018 Book of Why



Why? Three (Four) kinds of reasoning

BBNs support three main kinds of reasoning:

- **Predicting** conditions given predispositions
- **Diagnosing** conditions given symptoms (and predisposing)
- **Explaining** a condition by one or more predispositions

To which we can add a fourth:

- **Deciding** on an action based on probabilities of the conditions

Recall Bayes Rule

$$P(H,E) = P(H \mid E)P(E) = P(E \mid H)P(H)$$

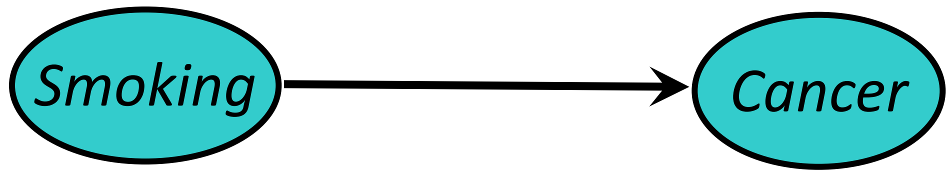
$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

$$P(E|H) = \frac{P(H|E) * P(E)}{P(H)}$$

Note symmetry: we can compute probability of a ***hypothesis given its evidence*** as well as probability of ***evidence given hypothesis***

Simple Bayesian Network

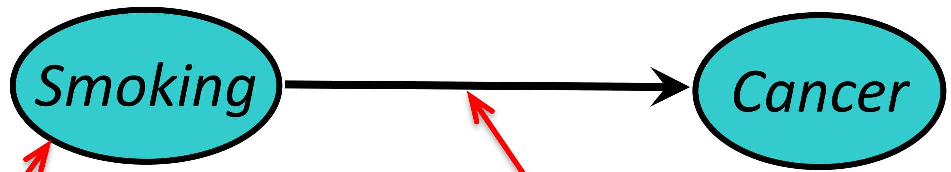
$$S \in \{no, light, heavy\}$$



$$C \in \{none, benign, malignant\}$$

Simple Bayesian Network

$$S \in \{no, light, heavy\}$$



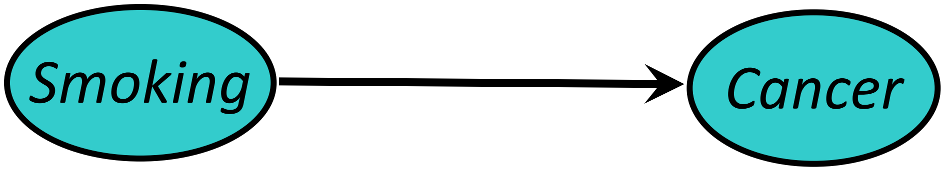
$$C \in \{none, benign, malignant\}$$

Nodes
represent
variables

Links represent
"causal" relations

Simple Bayesian Network

$S \in \{no, light, heavy\}$



$C \in \{none, benign, malignant\}$

Prior probability of S

P(S=no)	0.80
P(S=light)	0.15
P(S=heavy)	0.05

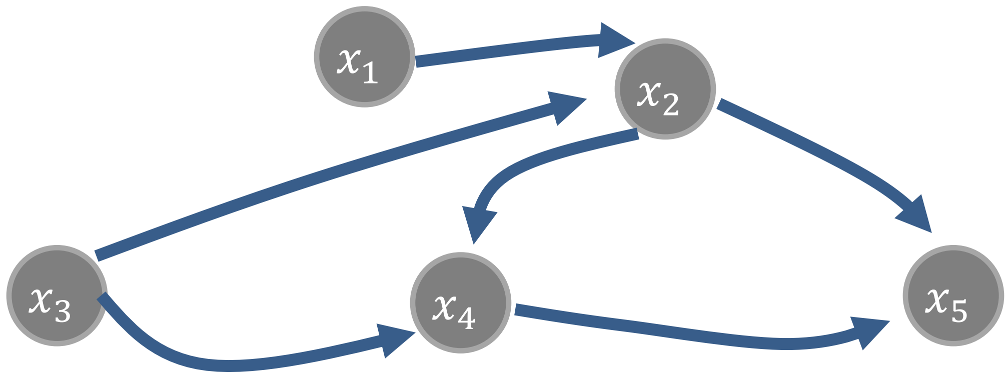
Nodes with no in-links have **prior probabilities**

Conditional distribution of S and C

Smoking=	*no*	*light*	*heavy*
C=none	0.96	0.88	0.60
C=benign	0.03	0.08	0.25
C=malignant	0.01	0.04	0.15

Nodes with in-links have **joint probability distributions**

Bayesian Networks: Directed Acyclic Graphs

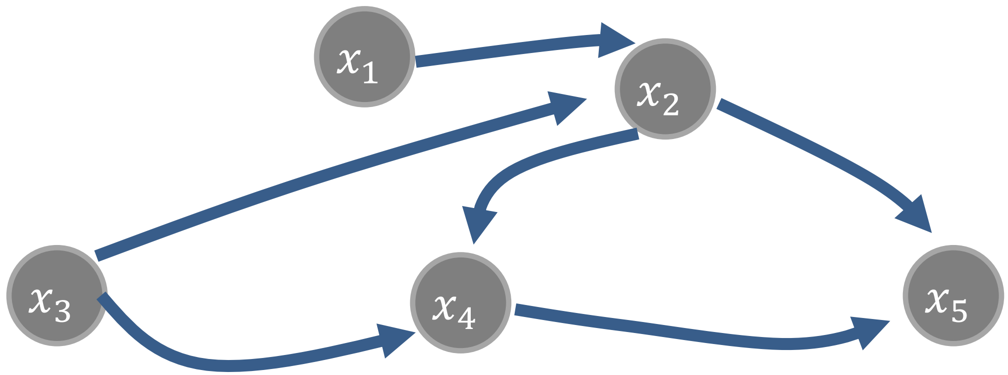


$$p(x_1, x_2, x_3, \dots, x_N) = \prod_i p(x_i \mid \pi(x_i))$$

topological sort

"parents of"

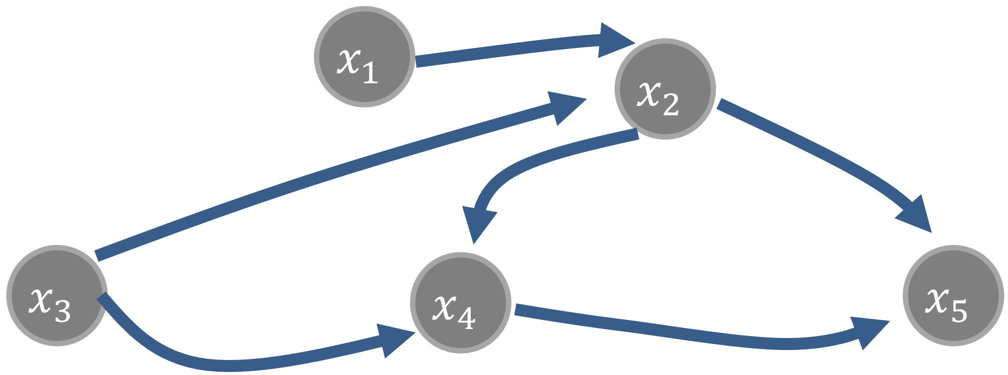
Bayesian Networks: Directed Acyclic Graphs



$$p(x_1, x_2, x_3, \ldots, x_N) = \prod_i p(x_i \mid \pi(x_i))$$

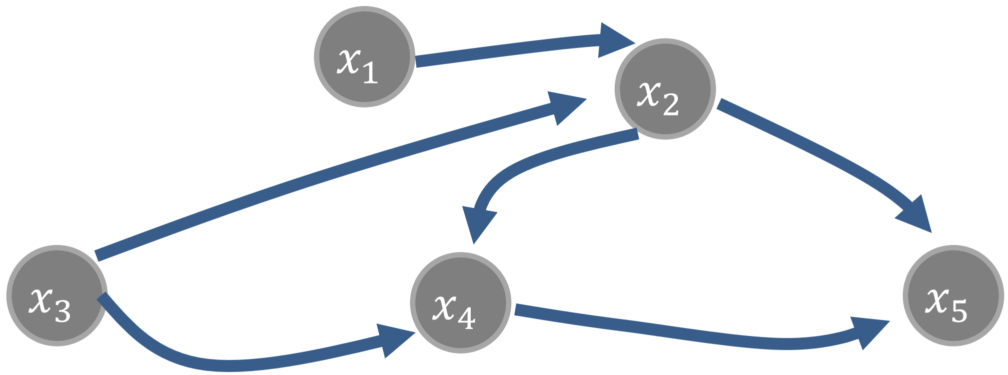
$$p(x_1, x_2, x_3, x_4, x_5) = ???$$

Bayesian Networks: Directed Acyclic Graphs



$$p(x_1, x_2, x_3, x_4, x_5) =$$
$$p(x_1)p(x_3)p(x_2|x_1, x_3)p(x_4|x_2, x_3)p(x_5|x_2, x_4)$$

Bayesian Networks: Directed Acyclic Graphs

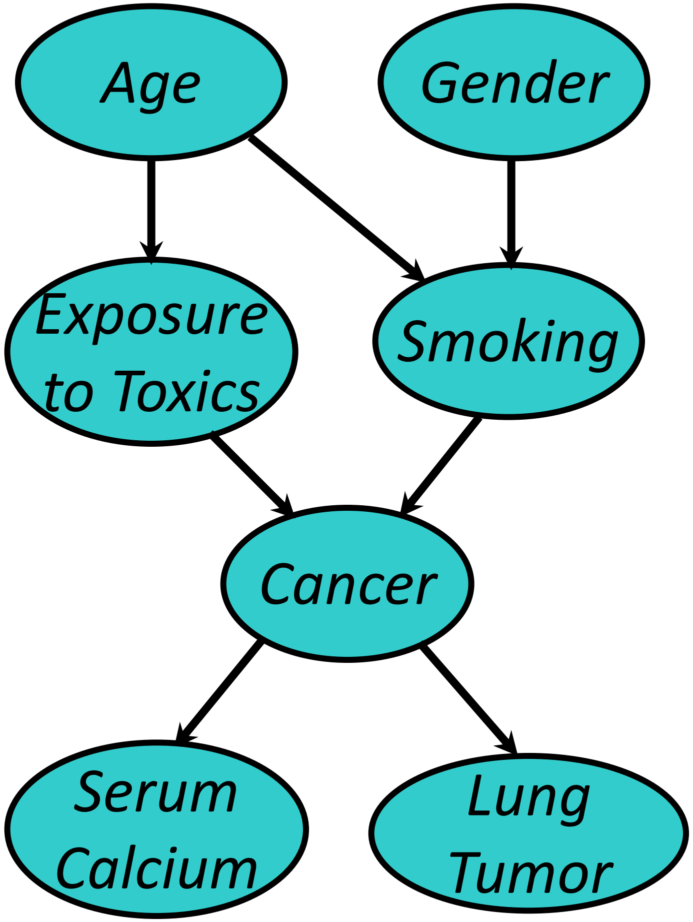


$$p(x_1, x_2, x_3, \ldots, x_N) = \prod_i p(x_i \mid \pi(x_i))$$

exact inference in general DAGs is NP-hard

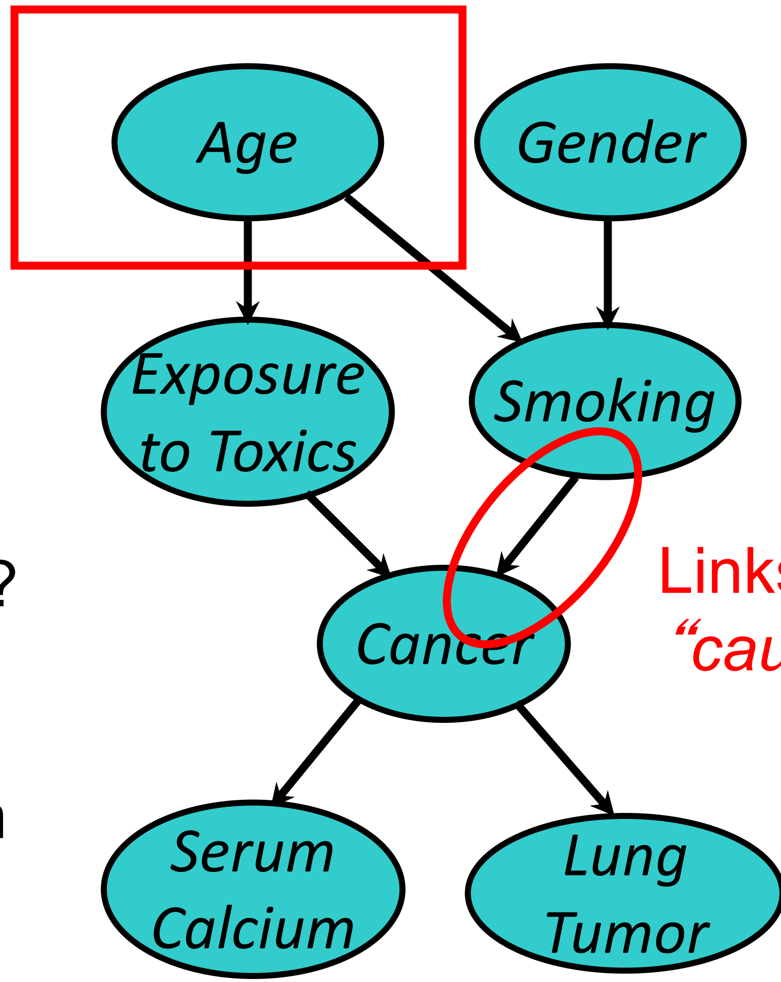
inference in trees can be exact

More Complex Bayesian Network



More Complex Bayesian Network

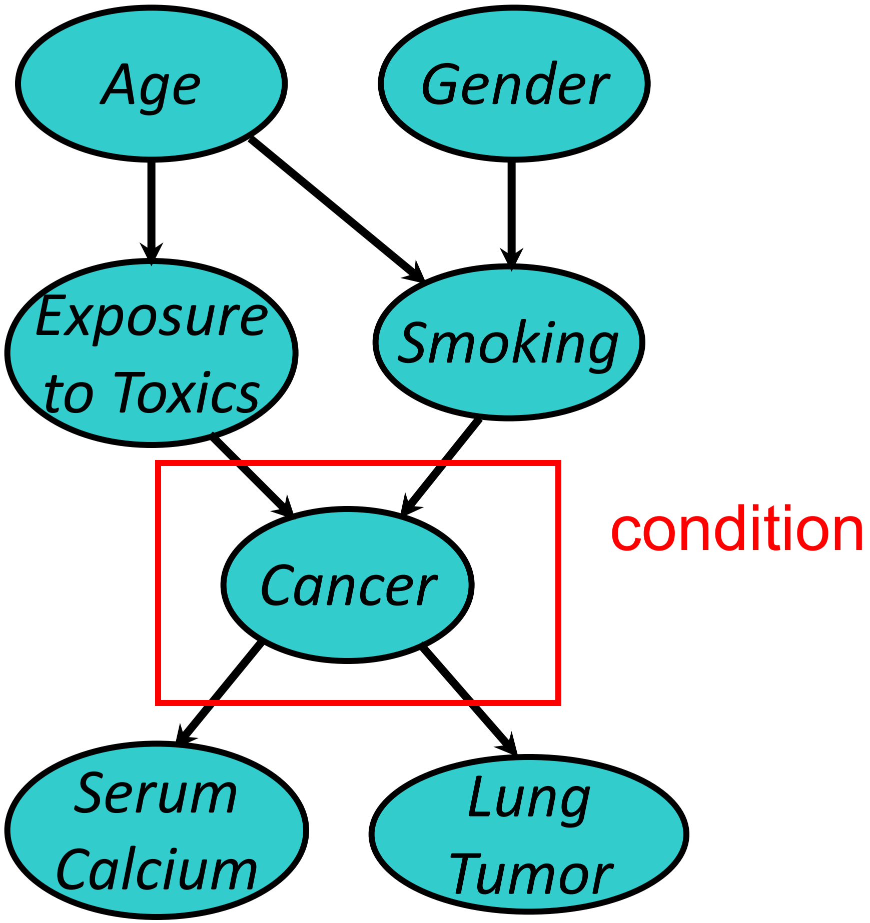
Nodes represent variables



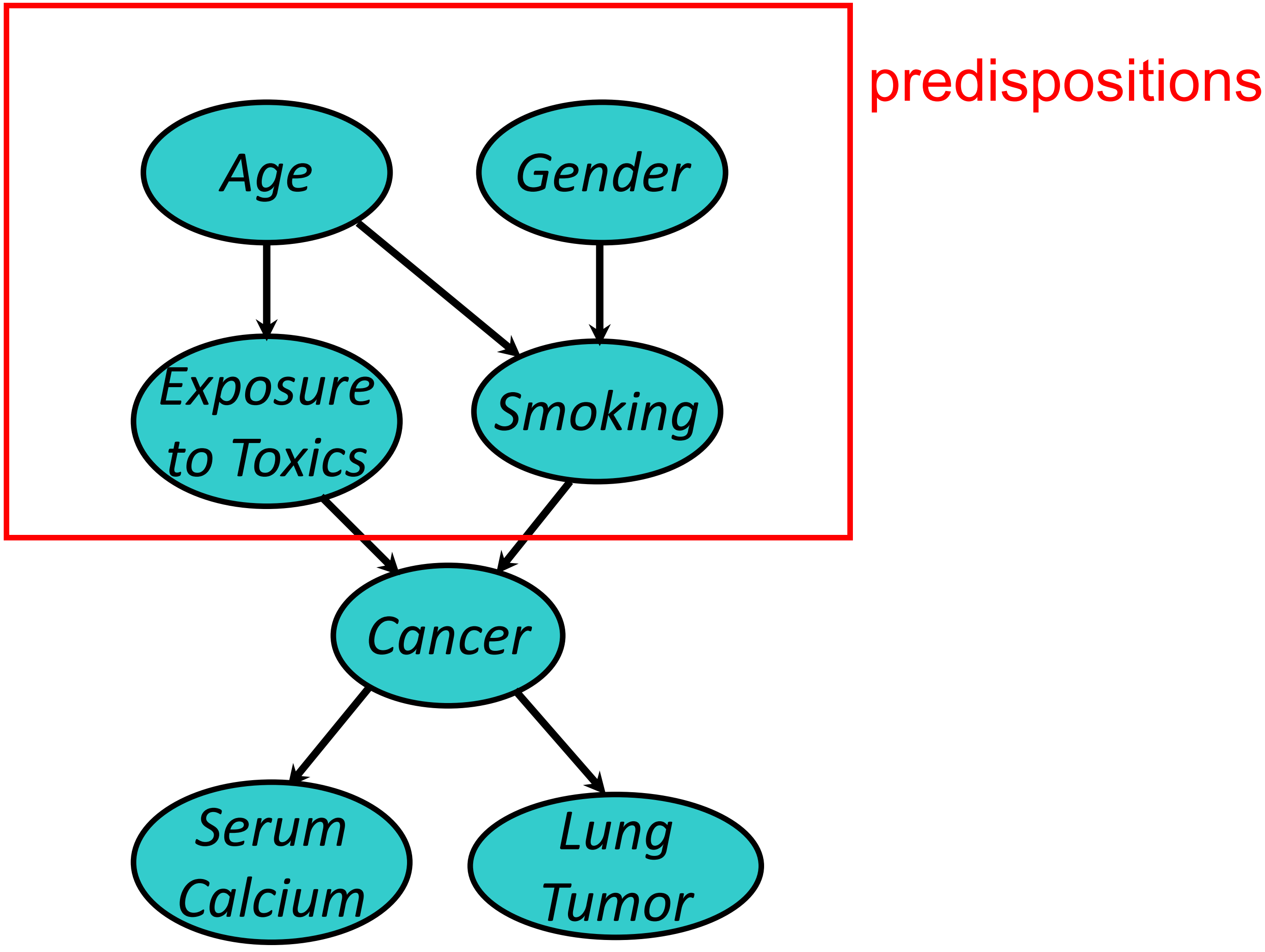
Links represent *"causal"* relations

- Does gender cause smoking?
- Influence might be a better term

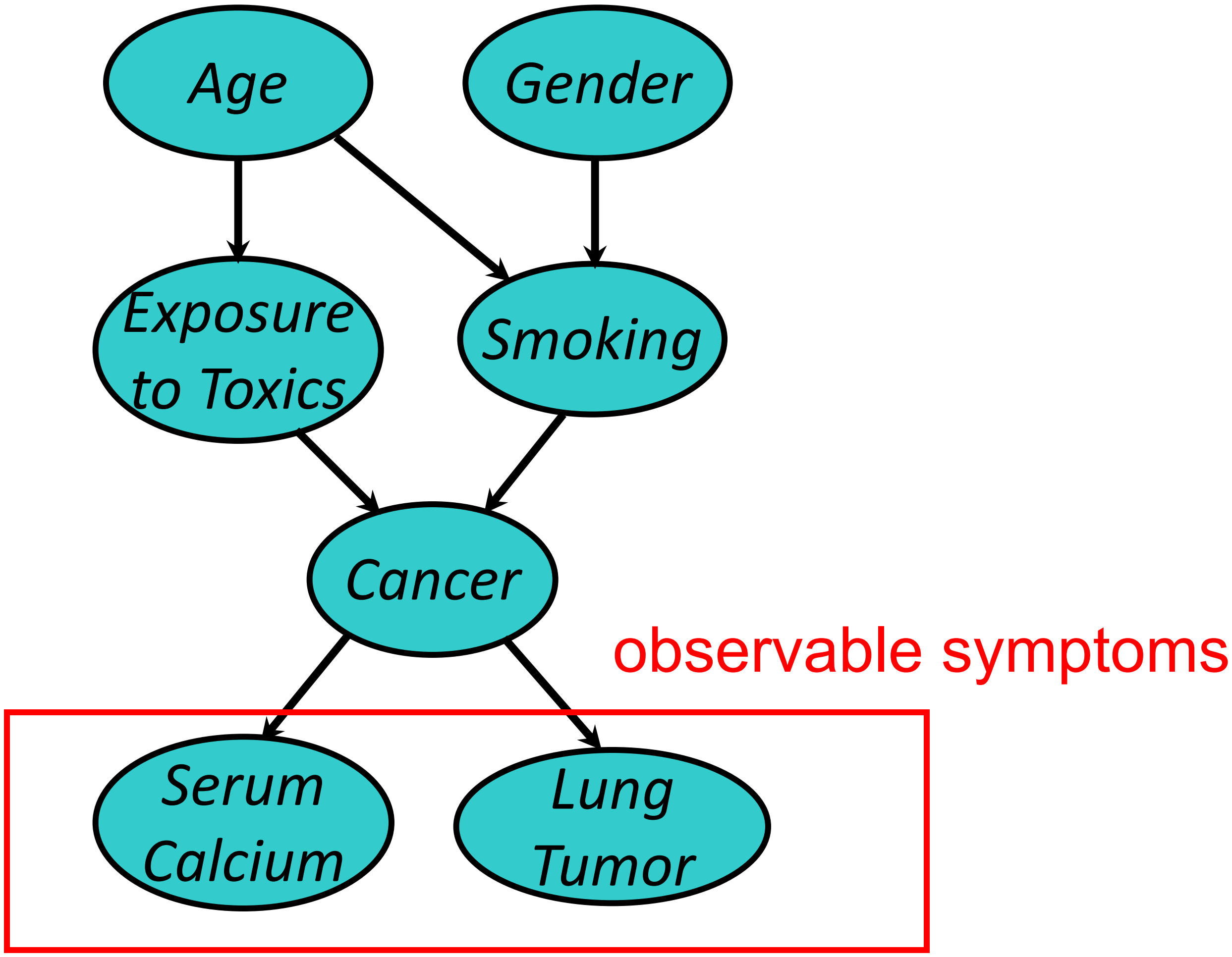
More Complex Bayesian Network



More Complex Bayesian Network

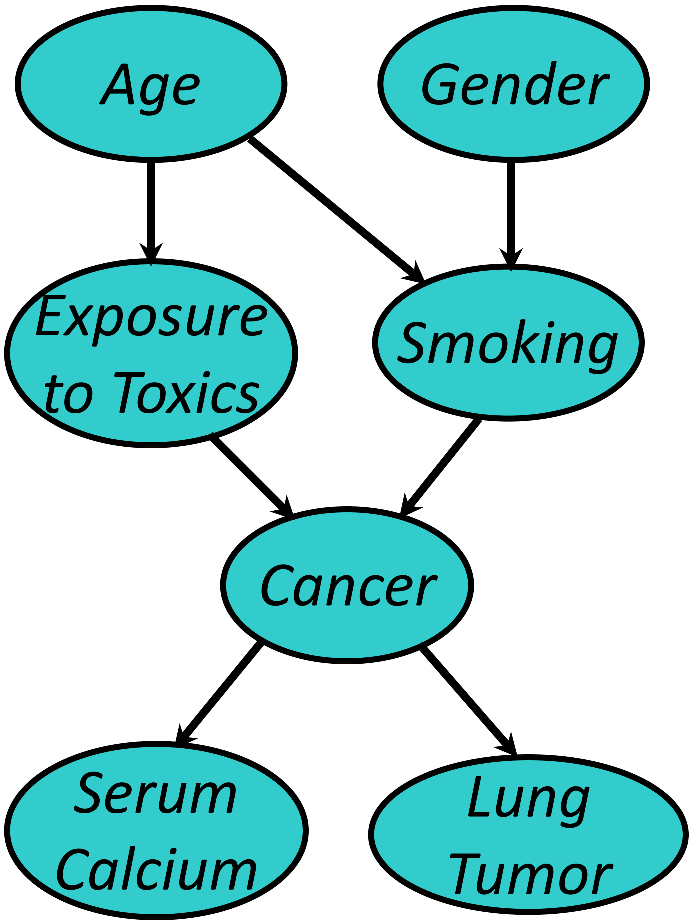


More Complex Bayesian Network



More Complex Bayesian Network

Can we predict likelihood of **lung tumor** given values of other 6 variables?



- Model has 7 variables
- Complete joint probability distribution will have 7 dimensions!
- Too much data required ☹
- BBN simplifies: a node has a CPT with data on itself & parents in graph

CPT = conditional probability table

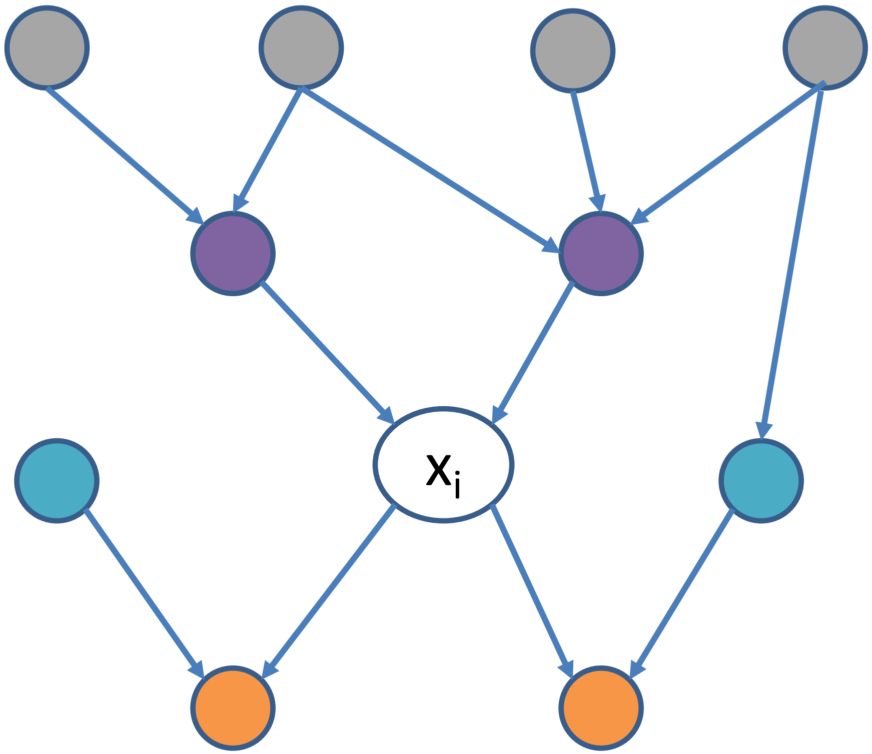
Independence & Conditional Independence in BBNs

Read these independence relationships right from the graph!

There are two common concepts that can help:

1. Markov blanket
2. D-separation (not covering)

Markov Blanket

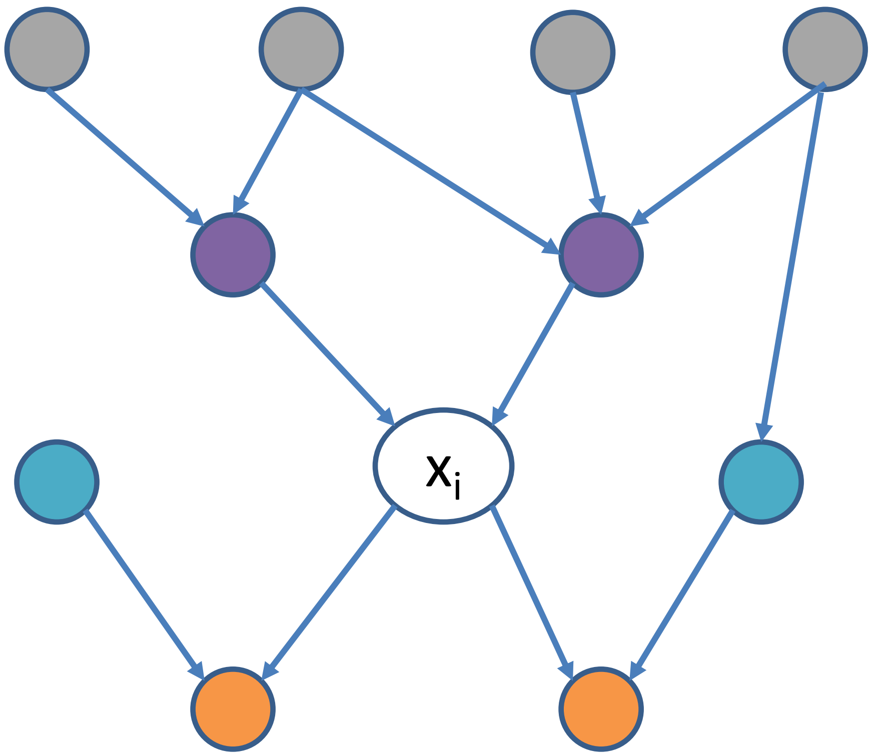


Markov blanket of a node x is its parents, children, and children's parents

(in this example, shading does not show observed/latent)

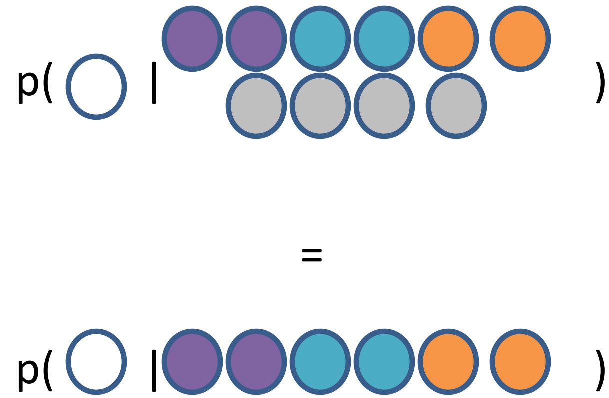
The **Markov Blanket** of a node x_i the set of nodes needed to form the complete conditional for a variable x_i

Markov Blanket



Markov blanket of a node x is its parents, children, and children's parents

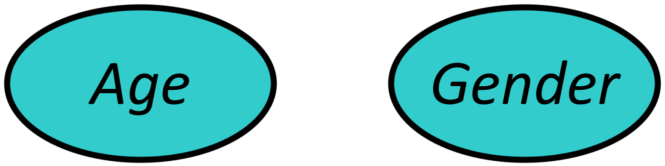
The **Markov Blanket** of a node x_i the set of nodes needed to form the complete conditional for a variable x_i



Given its Markov blanket, a node is conditionally independent of all other nodes in the BN

(in this example, shading does not show observed/latent)

Independence



Age and *Gender* are independent*.

$$P(A,G) = P(G) * P(A)$$

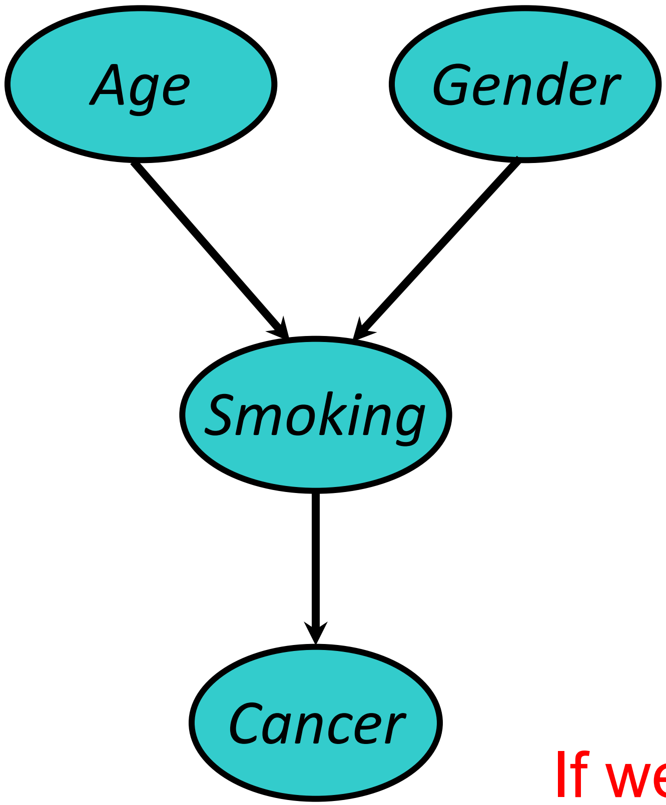
There is no path between them in the graph

$$P(A \mid G) = P(A)$$
$$P(G \mid A) = P(G)$$

$$P(A,G) = P(G \mid A)\, P(A) = P(G)P(A)$$
$$P(A,G) = P(A \mid G)\, P(G) = P(A)P(G)$$

* Not strictly true, but a reasonable approximation

Conditional Independence

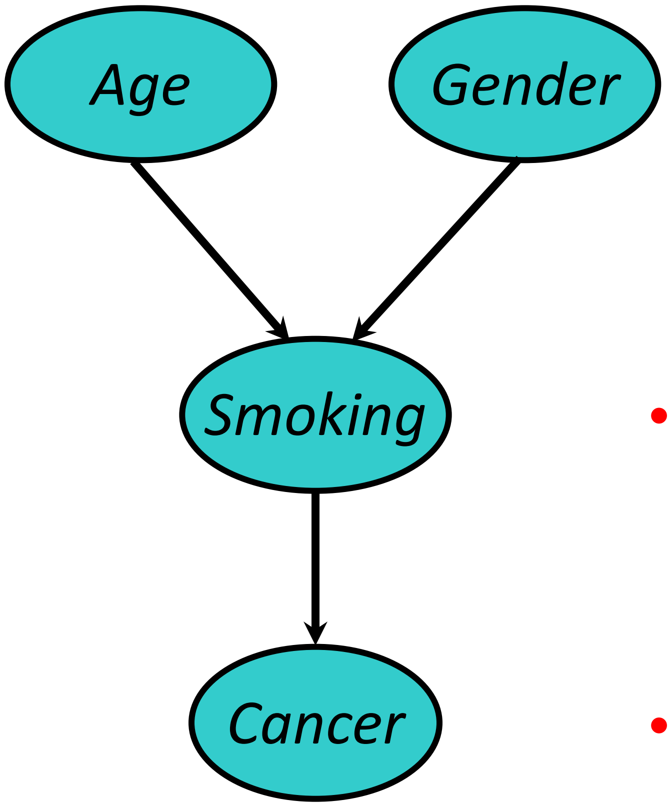


Cancer is independent of *Age* and *Gender* given *Smoking*

$$P(C \mid A,G,S) = P(C \mid S)$$

If we know value of smoking, no need to know values of age or gender

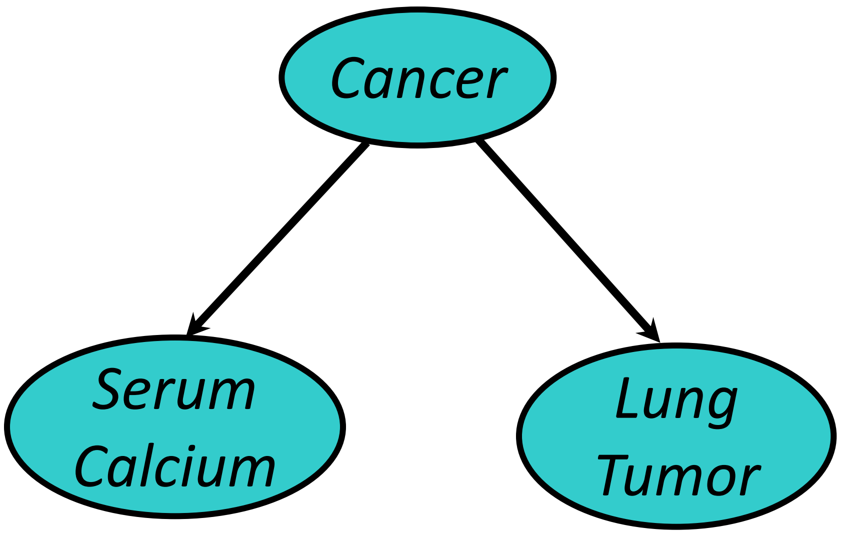
Conditional Independence



Cancer is independent of *Age* and *Gender* given *Smoking*

- Instead of one big CPT with 4 variables, we have two smaller CPTs with 3 and 2 variables
- If all variables binary: 12 models ($2^3 + 2^2$) rather than 16 (2^4)

Conditional Independence: Naïve Bayes



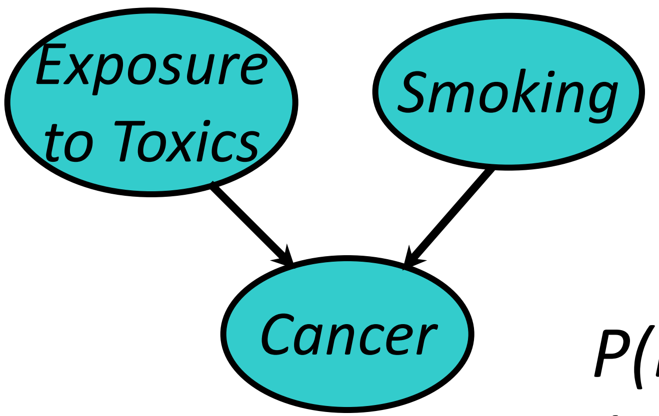
Serum Calcium and *Lung Tumor* are dependent

Serum Calcium is independent of *Lung Tumor*, given *Cancer*

$$P(L \mid SC, C) = P(L \mid C)$$
$$P(SC \mid L, C) = P(SC \mid C)$$

Naïve Bayes assumption: evidence (e.g., symptoms) independent given disease; easy to combine evidence

Explaining Away



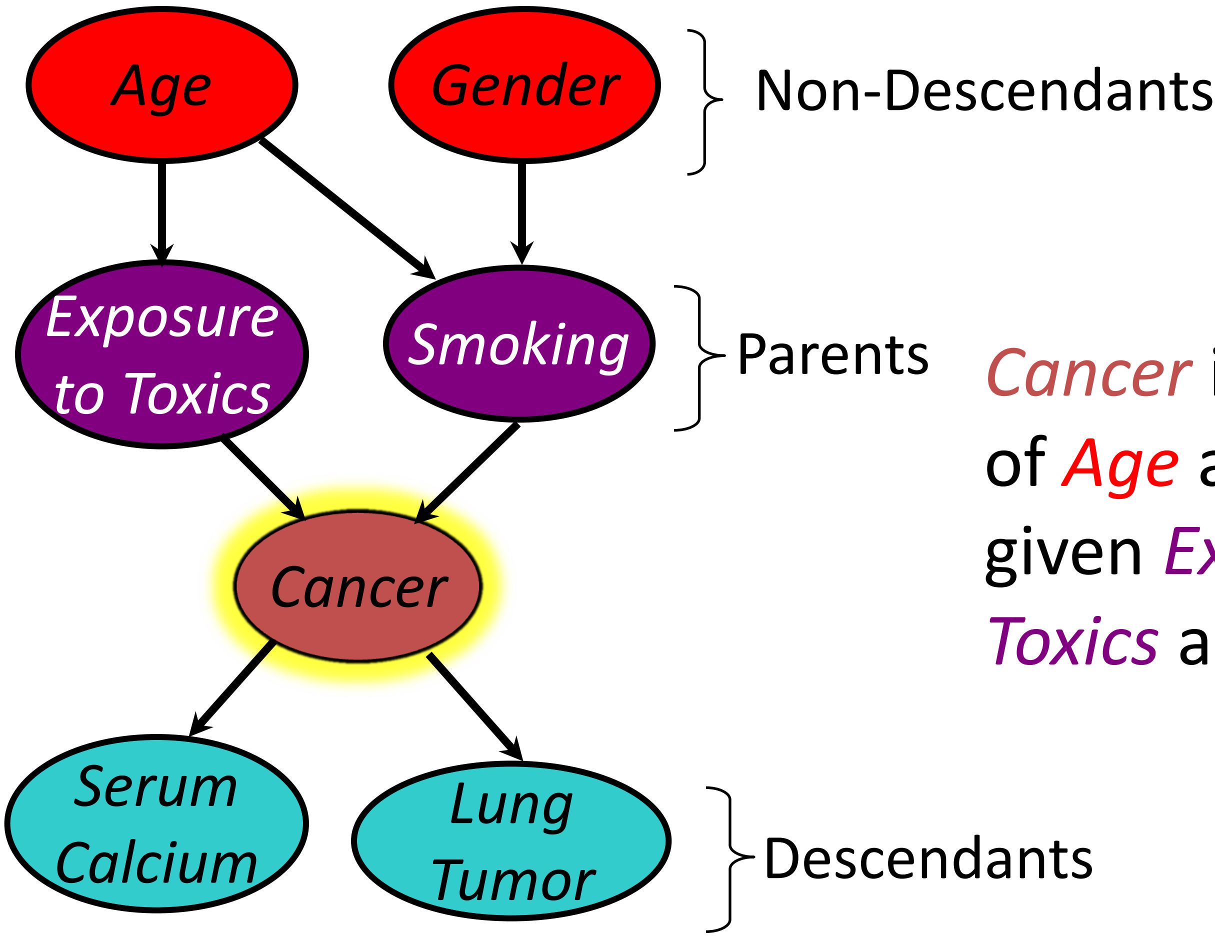
Exposure to Toxics and *Smoking* are independent

Exposure to Toxics is **dependent** on *Smoking*, given *Cancer*

P(E=heavy | C=malignant) > P(E=heavy | C=malignant, S=heavy)

- *Explaining away:* reasoning pattern where confirmation of one cause reduces need to invoke alternatives
- Essence of [Occam's Razor]() (prefer hypothesis with fewest assumptions)
- Relies on independence of causes

Conditional Independence



Cancer is independent of *Age* and *Gender* given *Exposure to Toxics* and *Smoking*.

BBN Construction

The **knowledge acquisition** process for a BBN involves three steps

KA1: Choosing appropriate variables

KA2: Deciding on the network structure

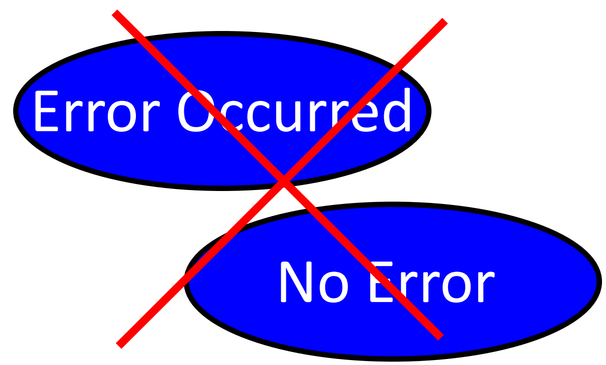
KA3: Obtaining data for the conditional probability tables

KA1: Choosing variables

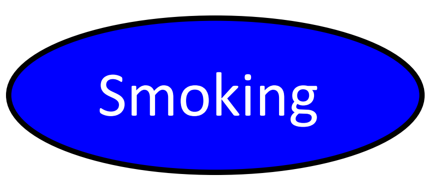
- Variable values: integers, reals or enumerations
- Variable should have collectively *exhaustive*, *mutually exclusive* values

$$x_1 \vee x_2 \vee x_3 \vee x_4$$

$$\neg(x_i \wedge x_j) \quad i \neq j$$



- They should be values, not probabilities

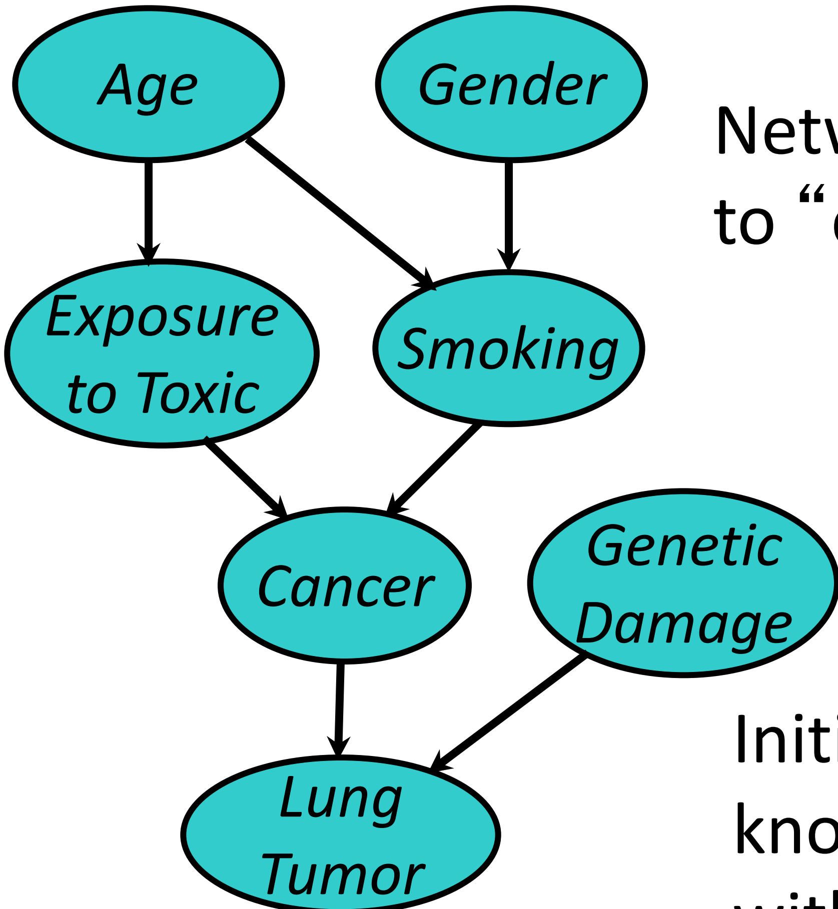


Heuristic: Knowable in Principle

Example of good variables

- Weather: {Sunny, Cloudy, Rain, Snow}
- Gasoline: Cents per gallon {0,1,2…}
- Temperature: { $\geq 100^\circ$ F , $< 100^\circ$ F}
- User needs help on Excel Charts: {Yes, No}
- User's personality: {dominant, submissive}

KA2: Structuring



Network structure corresponding to "causality" is usually good.

Initially this uses the designer's knowledge but can be checked with data

KA3: The Numbers

- For each variable we have a table of probability of its value for values of its **parents**
- For variables w/o parents, we have **prior probabilities**

$$S \in \{no, light, heavy\}$$

$$C \in \{none, benign, malignant\}$$

smoking priors	
no	0.80
light	0.15
heavy	0.05

	smoking		
cancer	**no**	**light**	**heavy**
none	0.96	0.88	0.60
benign	0.03	0.08	0.25
malignant	0.01	0.04	0.15

Three (Four) kinds of reasoning

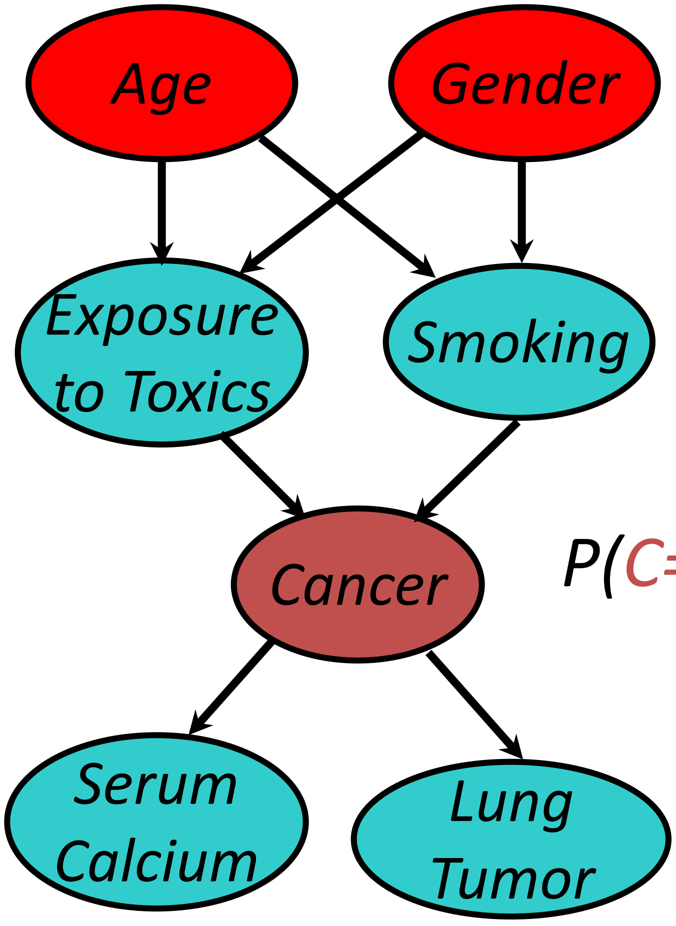
BBNs support three main kinds of reasoning:

- **Predicting** conditions given predispositions
- **Diagnosing** conditions given symptoms (and predisposing)
- **Explaining** a condition by one or more predispositions

To which we can add a fourth:

- **Deciding** on an action based on probabilities of the conditions

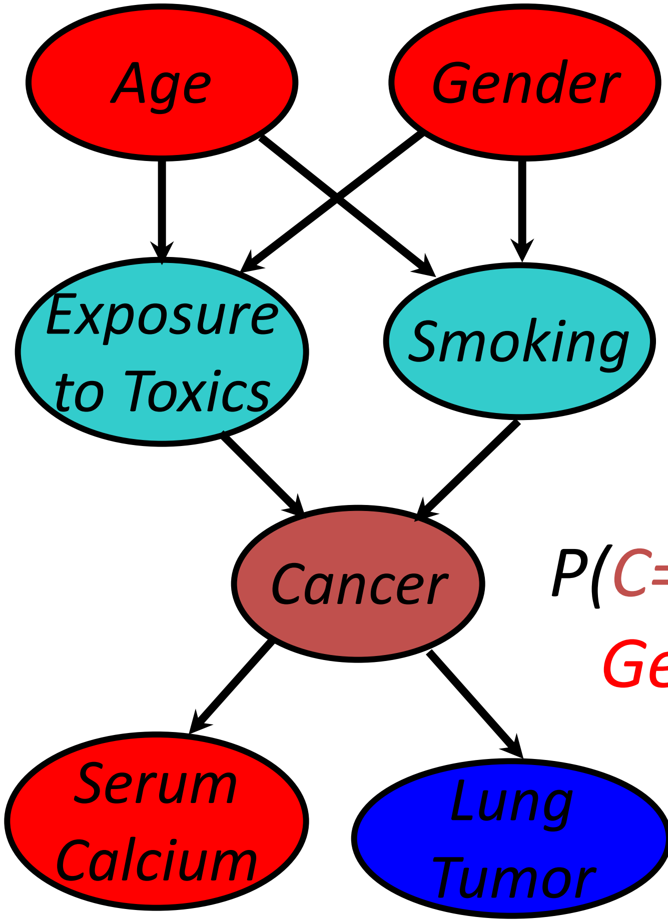
Predictive Inference



How likely are elderly males to get malignant cancer?

P(C=malignant | Age>60, Gender=male)

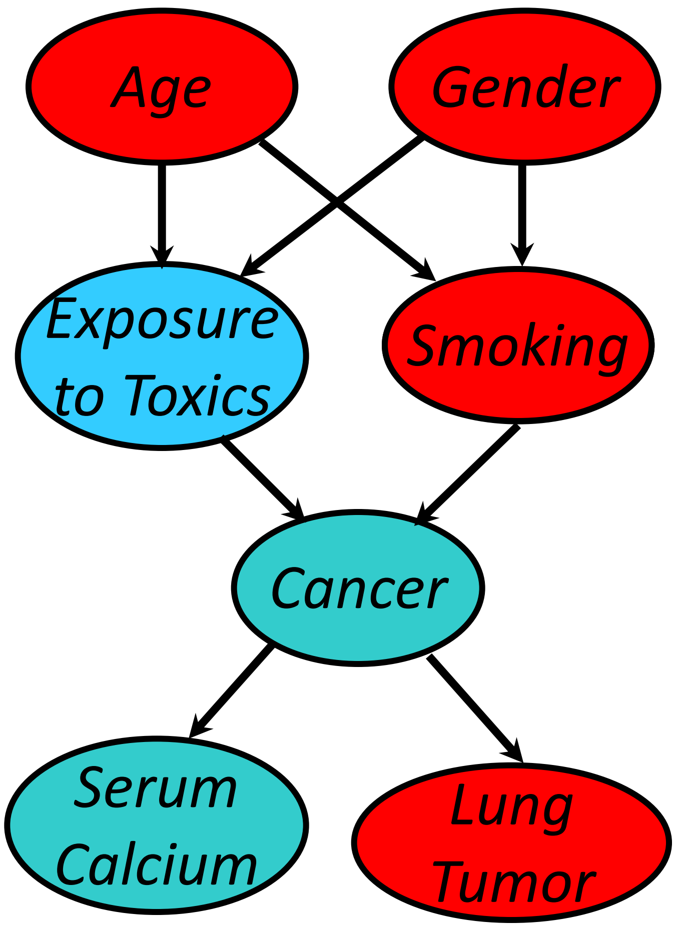
Predictive and diagnostic combined



How likely is an elderly male patient with high Serum Calcium to have malignant cancer?

P(C=malignant | Age>60, Gender= male, Serum Calcium = high)

Explaining away



- If we see a lung tumor, the probability of heavy smoking and of exposure to toxics both go up

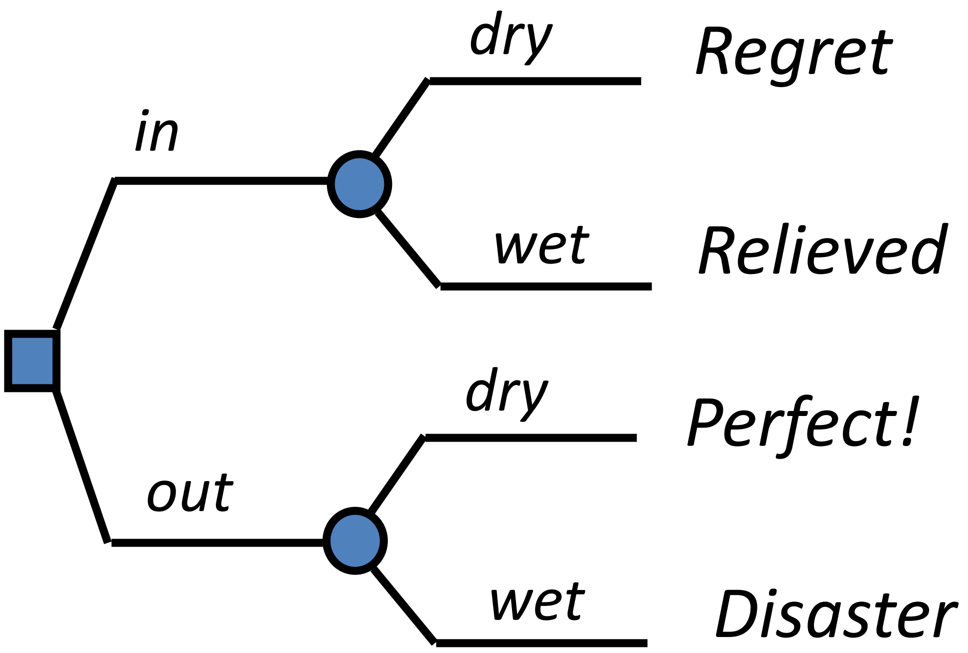
- If we then observe heavy smoking, the probability of exposure to toxics goes back down

Decision making

- A decision is a medical domain might be a choice of treatment (e.g., radiation or chemotherapy)
- Decisions should be made to **maximize expected utility**
- View decision making in terms of
  - Beliefs/Uncertainties
  - Alternatives/Decisions
  - Objectives/Utilities

Decision Problem

Should I have my party inside or outside?



Decision Making with BBNs

- Today's weather forecast might be either sunny, cloudy or rainy
- Should you take an umbrella when you leave?
- Your decision depends only on the forecast
  - The forecast "depends on" the actual weather
- Your satisfaction depends on your decision and the weather
  - Assign a utility to each of four situations: (rain|no rain) x (umbrella, no umbrella)

Decision Making with BBNs

- Extend BBN framework to include two new kinds of nodes: **decision** and **utility**
- **Decision** node computes the expected utility of a decision given its parent(s) (e.g., forecast) and a valuation
- **Utility** node computes utility value given its parents, e.g. a decision and weather
  - Assign utility to each situations: (rain|no rain) x (umbrella, no umbrella)
  - Utility value assigned to each is probably subjective

Fundamental Inference & Learning Question

- Compute posterior probability of a node given some other nodes

$$p(Q|x_1, \dots, x_j)$$

- Some techniques
  - MLE (maximum likelihood estimation)/MAP (maximum a posteriori) [covered 2nd]
  - Variable Elimination [covered 1st]
  - (Loopy) Belief Propagation ((Loopy) BP)
  - Monte Carlo
  - Variational methods
  - …

Advanced topics