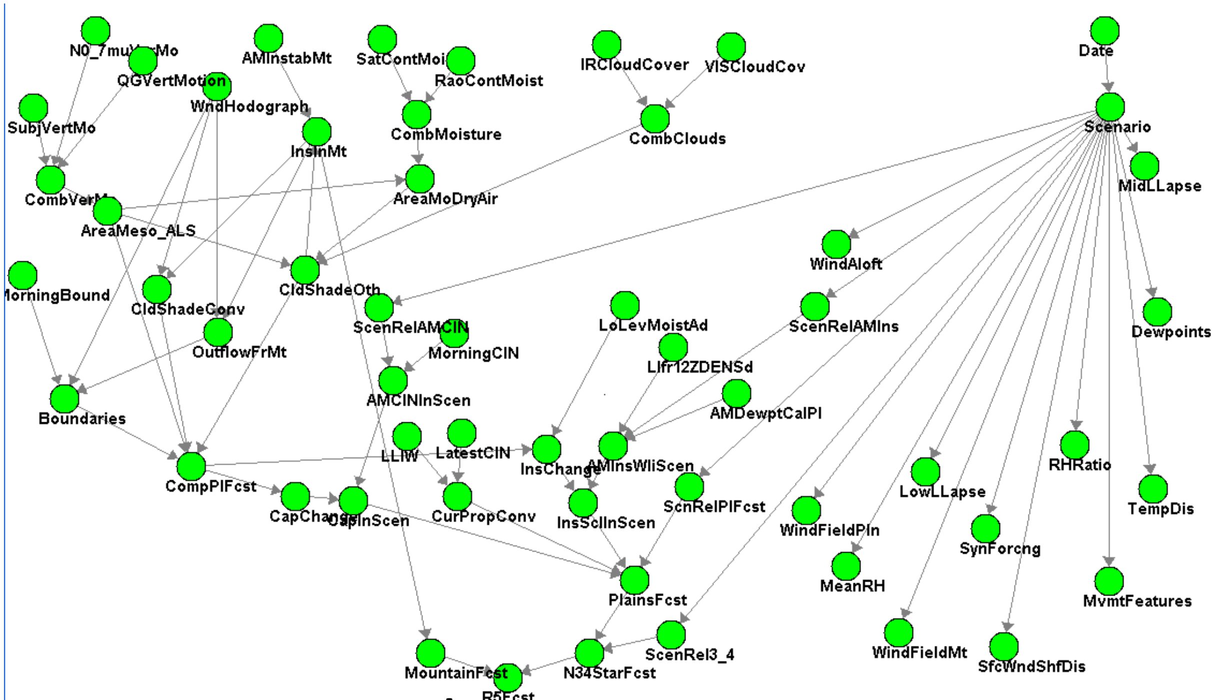
Probabilistic Graphical Models

Lecture 3 – Bayesian Networks Semantics

CS/CNS/EE 155
Andreas Krause

Bayesian networks

- Compact representation of distributions over large number of variables
- (Often) allows efficient exact inference (computing marginals, etc.)



JavaBayes applet

HailFinder

56 vars

~ 3 states each

➔ $\sim 10^{26}$ terms

> **10.000 years**

on Top supercomputers

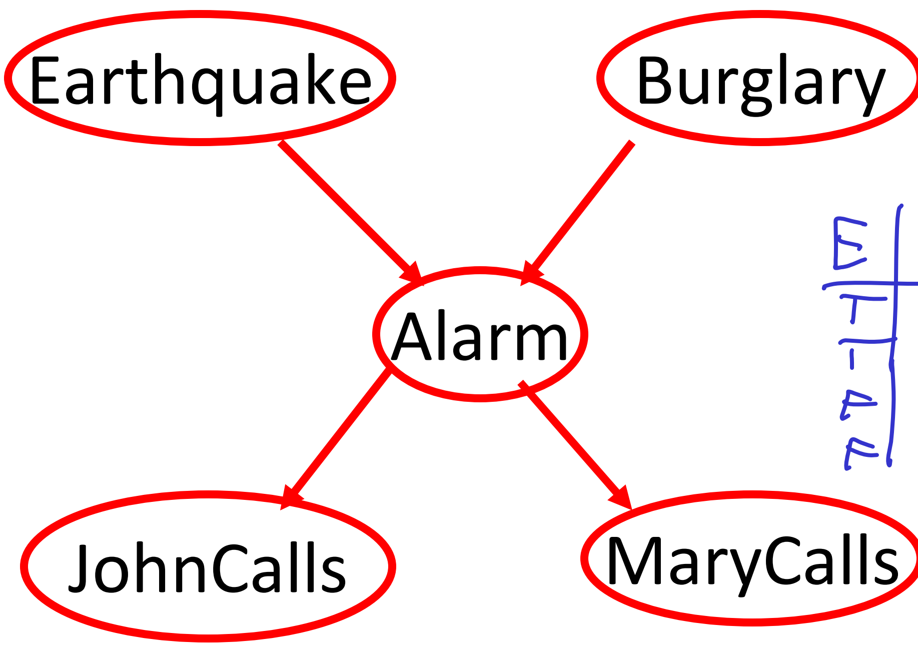
Causal parametrization

- Graph with directed edges from (immediate) causes to (immediate) effects

E	P(E)
T	.01
F	.99

B	P(B)
T	.03
F	.97

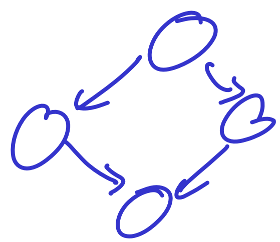
$P(E|J)$



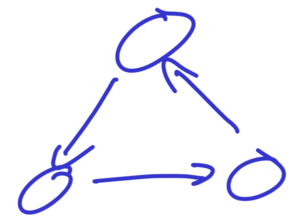
E	B	P(A|E,B)
T	T	.
T	F	.
F	T	.
F	F	.

Bayesian networks

- A **Bayesian network structure** is a directed, acyclic graph G, where each vertex s of G is interpreted as a random variable X_s (with unspecified distribution)



OK



NOT OK

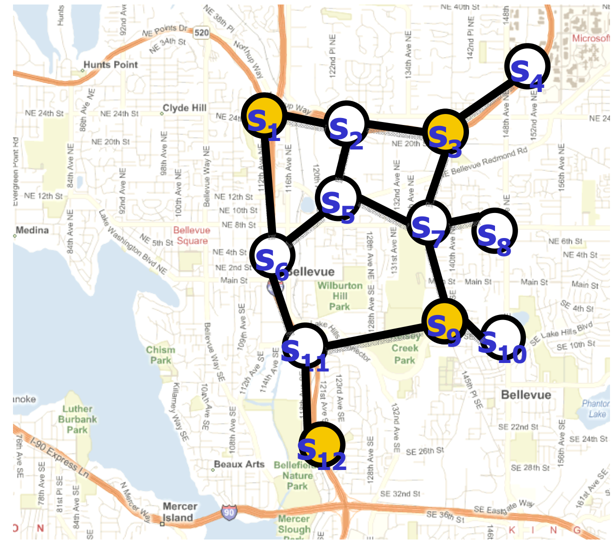
- A **Bayesian network** (G,P) consists of
  - A BN structure G and ..
  - ..a set of conditional probability distributions (CPDs) $P(X_s \mid \mathbf{Pa}_{X_s})$, where $\mathbf{Pa}_{X_s}$ are the parents of node X_s such that
  - (G,P) defines joint distribution

$$P(X_1, ..., X_n) = \prod_i P(X_i \mid \mathbf{Pa}_{X_i})$$

Representing the world using BNs

True distribution P’
with cond. ind. I(P’)

represent



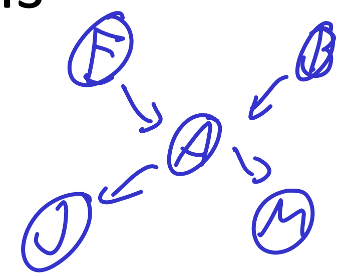
Bayes net (G,P)
with I(P)

- Want to make sure that I(P) $\subseteq$ I(P’)
- Need to understand CI properties of BN (G,P)

Local Markov Assumption

- Each BN Structure G is associated with the following conditional independence assumptions

$J \perp B \mid A$

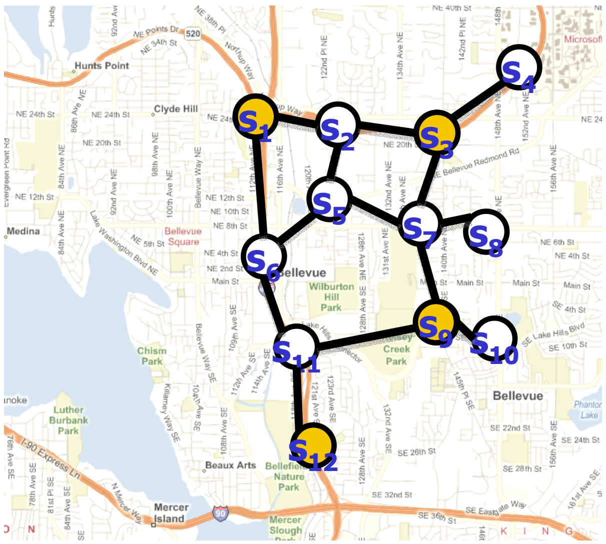


$$X \perp \textbf{NonDescendents}_X \mid \textbf{Pa}_X$$

- We write $I_{loc}(G)$ for these conditional independences

- Suppose (G,P) is a Bayesian network representing P
  Does it hold that $I_{loc}(G) \subseteq I(P)$?
  If this holds, we say G is an I-map for P.

Factorization Theorem



$I_{loc}(G) \subseteq I(P)$

G is an **I-map** of P
(independence map)

True distribution P
can be represented exactly as
Bayesian network (G,P)

$$P(X_1, ..., X_n) = \prod_i P(X_i \mid \mathbf{Pa}_{X_i})$$

Additional conditional independencies

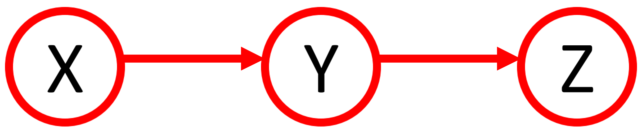
- BN specifies joint distribution through conditional parameterization that satisfies Local Markov Property $I_{loc}(G) = \{(X_i \perp \text{Nondescendants}_{X_i} \mid \text{Pa}_{X_i})\}$
- But we also talked about additional properties of CI
  - Weak Union, Intersection, Contraction, …
- Which additional CI does a particular BN specify?
  - All CI that can be derived through algebraic operations

➔ proving CI is very cumbersome!!

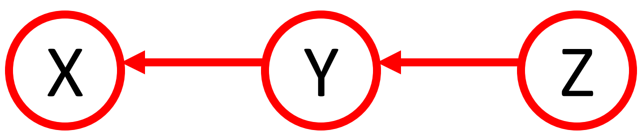
Is there an easy way to find all independences of a BN just by looking at its graph??

BNs with 3 nodes

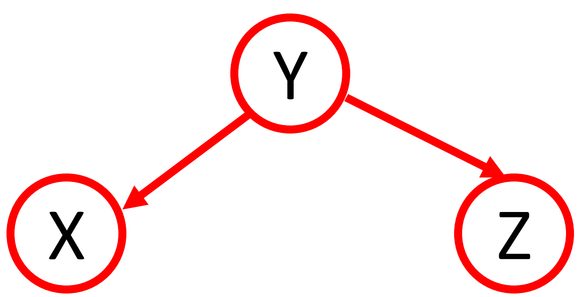
Local Markov Property:
$X \perp \mathrm{NonDesc}(X) \mid \mathrm{Pa}(X)$



$X \perp Z \mid Y$
$\neg(X \perp Z)$

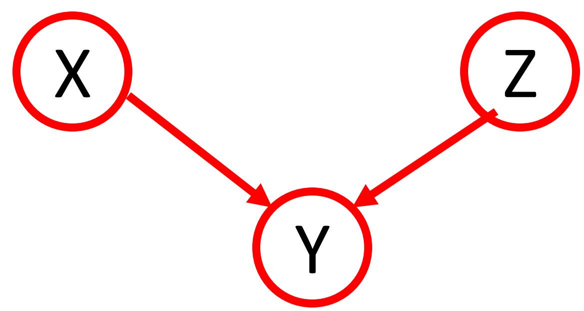


$X \perp Z \mid Y$
$\neg(X \perp Z)$

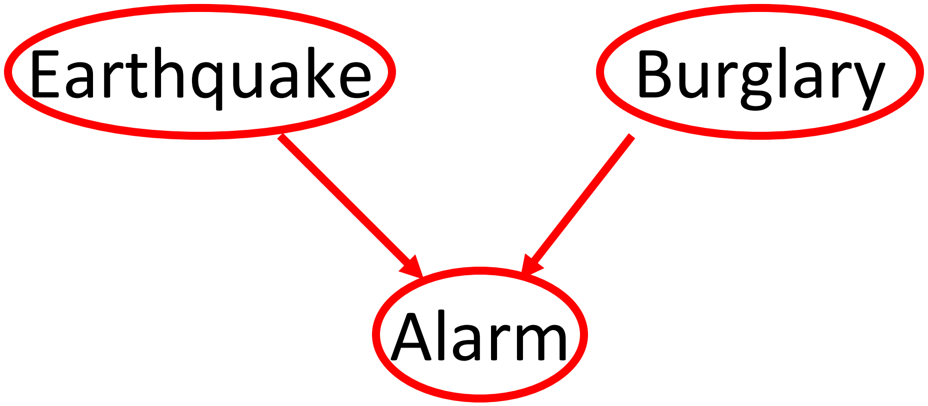


$X \perp Z \mid Y$
$\neg(X \perp Z)$

$X \perp Z$
$\neg(X \perp Z \mid Y)$



V-structures



- Know E $\perp$ B
- Suppose we know A. Does E $\perp$ B | A hold?

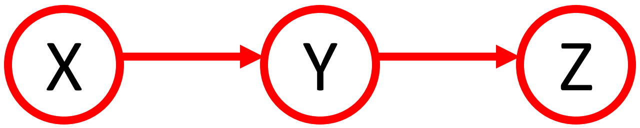
Can happen: $P(E=T \mid A=T, B=T) < P(E=T \mid A=T)$

Explaining away

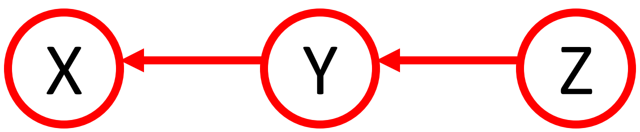
BNs with 3 nodes

Local Markov Property:
$X \perp \mathrm{NonDesc}(X) \mid \mathrm{Pa}(X)$

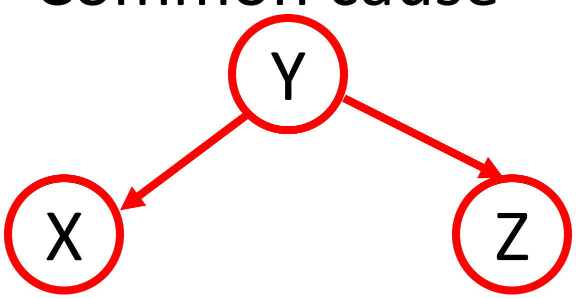
Indirect causal effect



Indirect evidential effect



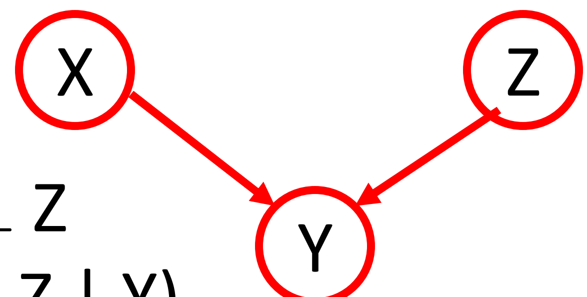
Common cause



$X \perp Z \mid Y$

$\neg\ (X \perp Z)$

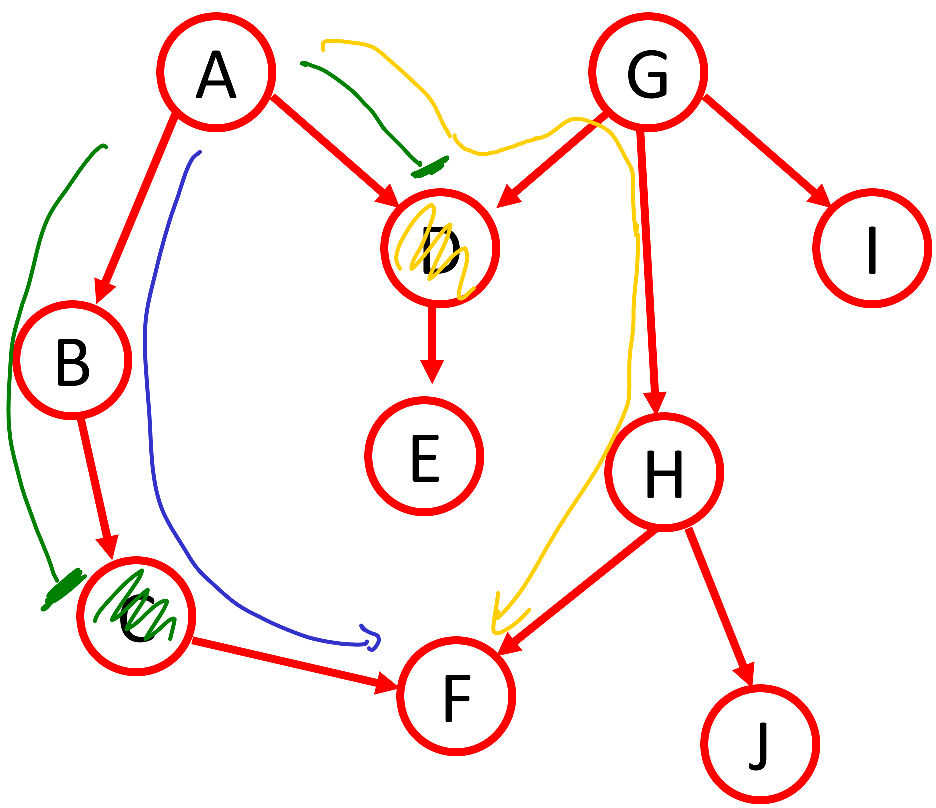
Common effect



$X \perp Z$

$\neg\ (X \perp Z \mid Y)$

Examples

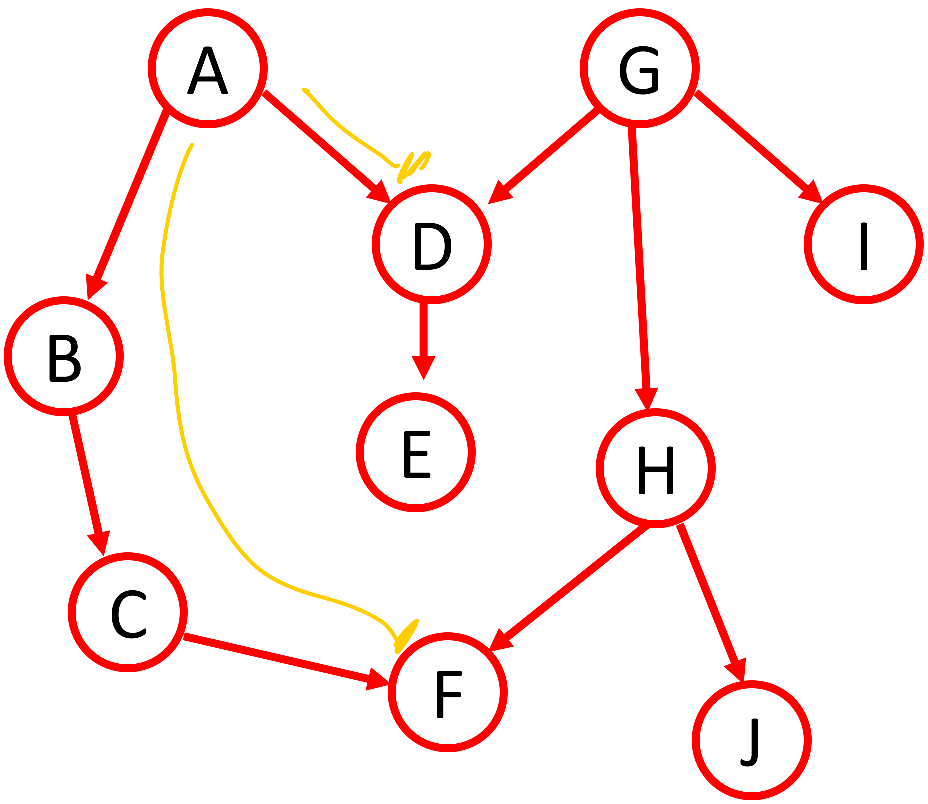


$A \perp F$ ✓

$A \perp F \mid C$ ✓

$A \perp F \mid C, D$ ✗

More examples



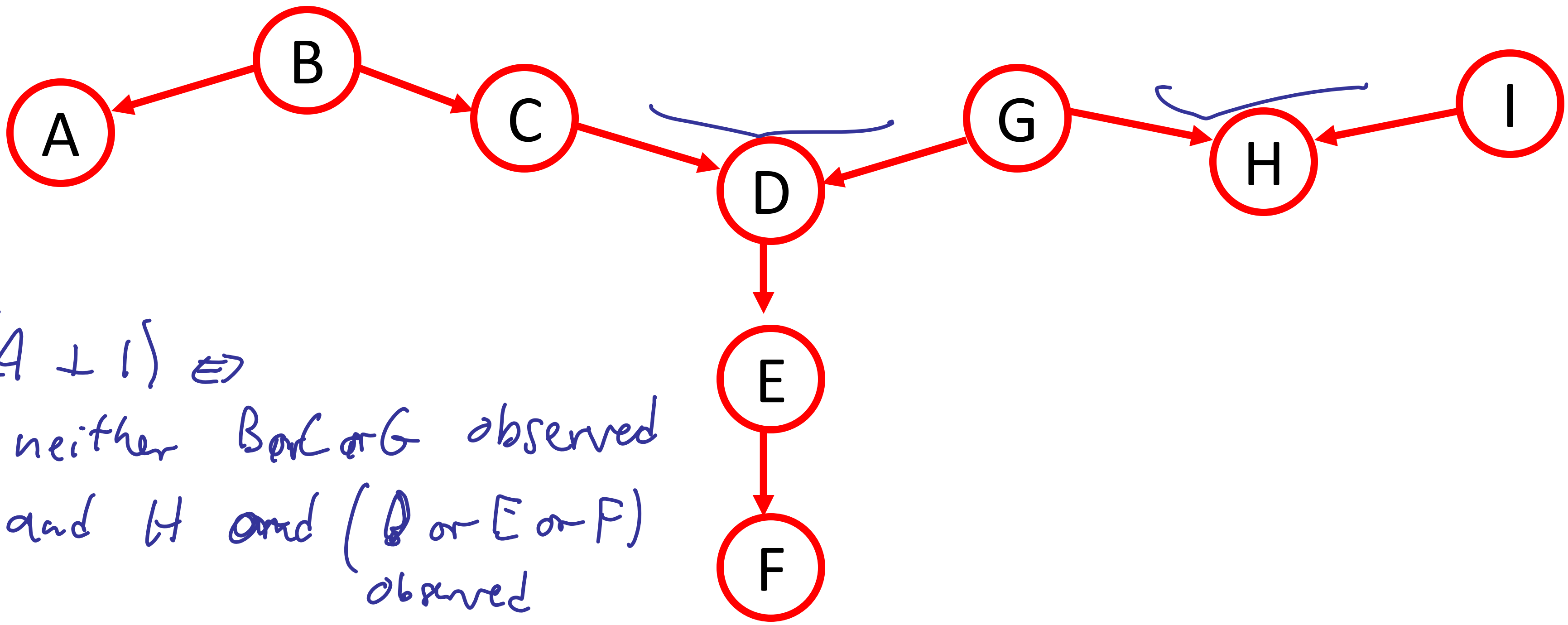
$A \perp G$

$A \perp G \mid D$ ×

$A \perp G \mid E$

Active trails

- When are A and I independent?



$\neg(A \perp I) \Leftrightarrow$
neither B or C or G observed
and H and (D or E or F)
observed

Active trails

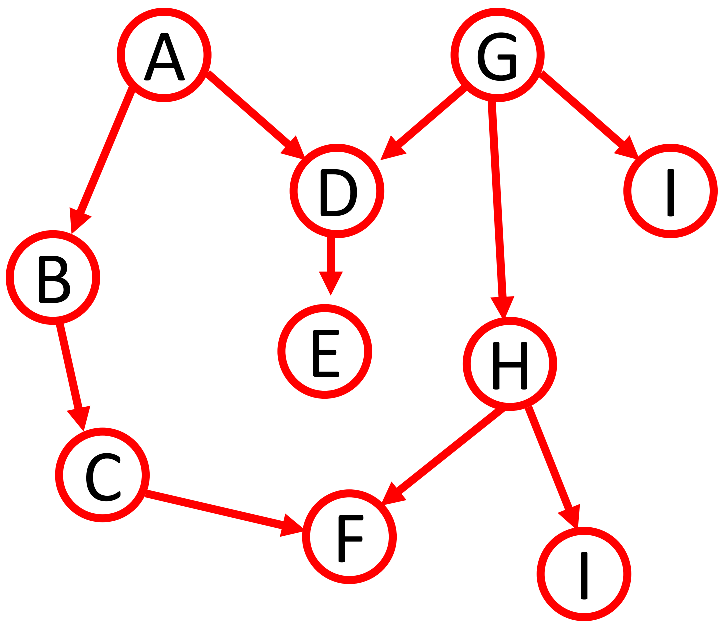
- An undirected path in BN structure G is called **active trail** for observed variables $\mathbf{O} \subseteq \{X_1,\ldots,X_n\}$, if for every consecutive triple of vars X,Y,Z on the path

blocked if $Y \in O$

  - $X \rightarrow Y \rightarrow Z$ and Y is unobserved ($Y \notin \mathbf{O}$)
  - $X \leftarrow Y \leftarrow Z$ and Y is unobserved ($Y \notin \mathbf{O}$)
  - $X \leftarrow Y \rightarrow Z$ and Y is unobserved ($Y \notin \mathbf{O}$)
  - $X \rightarrow Y \leftarrow Z$ and Y **or any of Y's descendants** is observed

- Any variables X_i and X_j for which $\nexists$ active trail for observations **O** are called d-separated by **O**
  We write **d-sep(X_i;X_j | O)**
- Sets **A** and **B** are d-separated given **O** if d-sep(X,Y |**O**) for all $X \in \mathbf{A}$, $Y \in \mathbf{B}$. Write **d-sep(A; B | O)**

d-separation and independence



Theorem:

d-sep(X;Y | **Z**) ➔ X ⊥ Y | **Z**

i.e., X cond. ind. Y given Z
if there does not exist
any active trail
between X and Y
for observations **Z**

- Proof uses algebraic properties of conditional independence

Soundness of d-separation

- Have seen: P factorizes according to G $\Leftrightarrow$ $I_{loc}(G) \subseteq I(P)$
- Define $I(G) = \{(X \perp Y \mid Z): \text{d-sep}_G(X;Y \mid Z)\}$

- **Theorem**: Soundness of d-separation

  P factorizes over G $\Rightarrow$ $I(G) \subseteq I(P)$

- Hence, d-separation captures only true independences

- How about $I(G) = I(P)$?

Does the converse hold?

Suppose P factorizes over G.

Does it hold that I(P) $\subseteq$ I(G)?

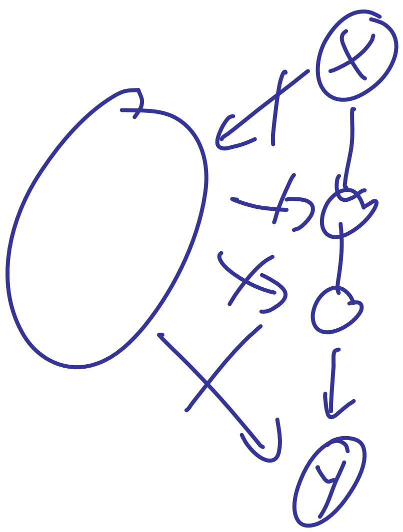
$P \models X \perp Y \qquad I(P) = \{(X \perp Y)\}$

$G: X \to Y \qquad I(G) = \{\}$

Existence of dependences for non-d-separated variables

- **Theorem**: If X and Y are not d-separated given **Z**, then there exists some distribution P factorizing over G in which X and Y are dependent given **Z**

- **Proof sketch**:



Pick active trail
Parameterize CPDs along trail to create dependence
Everything else set to independent to avoid cancelling dependencies

Completeness of d-separation

- Theorem: For "almost all" distributions P that factorize over G it holds that **I(G) = I(P)**
  - "almost all": except for a set of distributions with measure 0, assuming only that no finite set of distributions has measure > 0

$X \longrightarrow Y$

$P(X=T) = p$

$P(Y=T \mid X=T) = r$

$P(Y=T \mid X=F) = q$

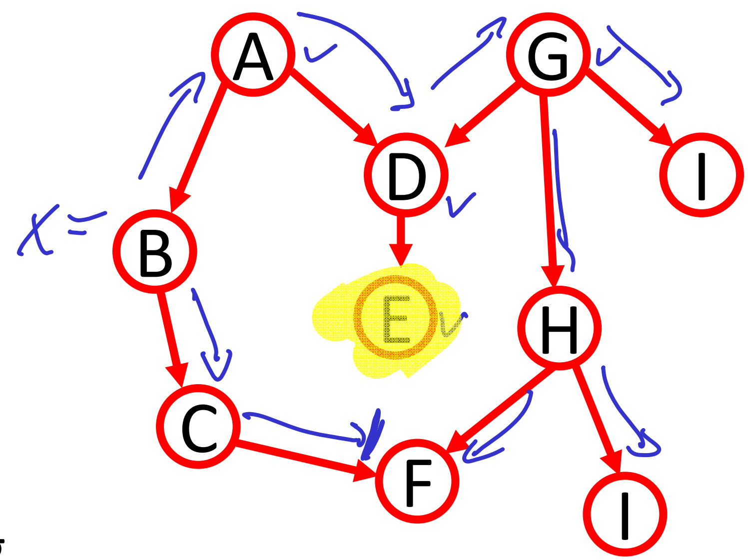
$P(Y \mid X) = P(Y)$

$$\underbrace{P(Y=T \mid X=T)}_{r} \overset{!}{=} \underbrace{P(Y=T)}_{rp + q(1-p)}$$

$$\underbrace{r \overset{!}{=} rp + q(1-p) \Rightarrow r(1-p) = q(1-p)}_{\text{happens with prob. } 0}$$

Algorithm for d-separation

- How can we check if X ⊥ Y | **Z**?
  - Idea: Check every possible path connecting X and Y and verify conditions
  - Exponentially many paths!!! ☹

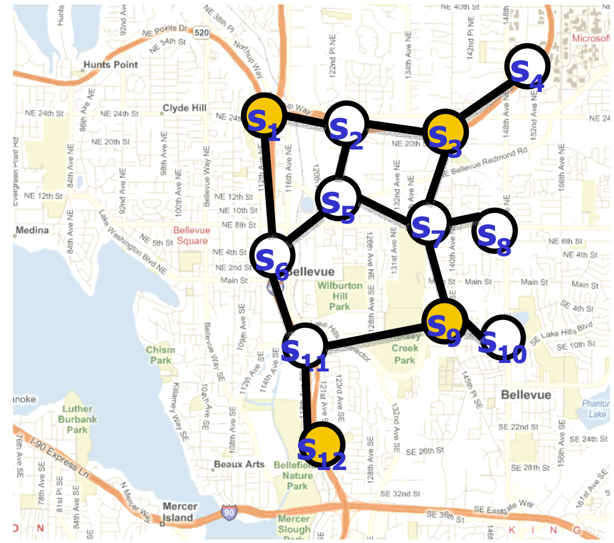


- Linear time algorithm:
  Find all nodes reachable from X
  - 1. Mark **Z** and its ancestors
  - 2. Do breadth-first search starting from X; stop if path is blocked
  - Have to be careful with implementation details (see reading)

Representing the world using BNs

True distribution P'
with cond. ind. I(P')

represent



Bayes net (G,P)
with I(P)

- Want to make sure that I(P) $\subseteq$ I(P')
- Ideally: I(P) = I(P')
- Want BN that **exactly** captures independencies in P'!

Minimal I-maps

- Lemma: Suppose G' is derived from G by adding edges
- Then $I(G') \subseteq I(G)$
- Proof:

$$I_{loc}(G') \subseteq I_{loc}(G)$$

Completeness: $I(G) = \{$ all CIs derivable from $I_{loc}(G)$ using CI properties $\}$

$$\Rightarrow I(G') \subseteq I(G)$$

- Thus, want to find graph G with $I(G) \subseteq I(P)$ such that when we remove any single edge, for the resulting graph G' it holds that $I(G') \not\subseteq I(P)$
- Such a graph G is called **minimal I-map**

Existence of Minimal I-Maps

- Does every distribution have a minimal I-Map?

Yes: Start with full graph G, $I(G) = \emptyset$

Keep removing edges as long as

$I(G) \subseteq I(P)$

Algorithm for finding minimal I-map

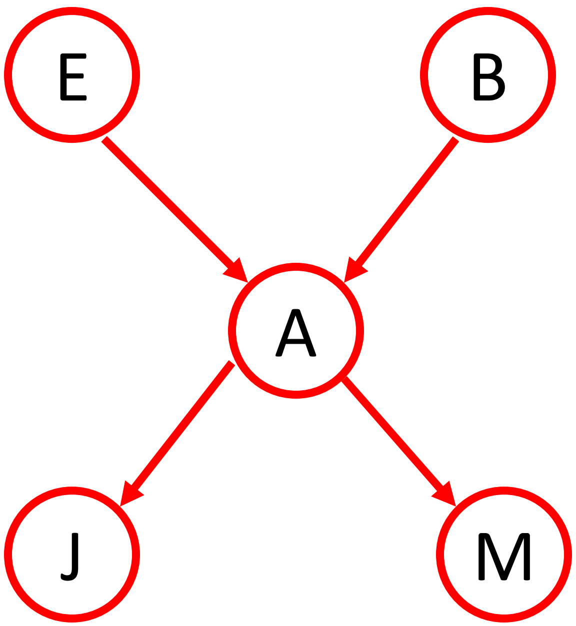
- Given random variables and known conditional independences
- Pick ordering $X_1,\ldots,X_n$ of the variables
- For each X_i
  - Find minimal subset $\mathbf{A} \subseteq \{X_1,\ldots,X_{i-1}\}$ such that $P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid \mathbf{A})$
  - Specify / learn CPD $P(X_i \mid \mathbf{A})$
- Will produce minimal I-map!

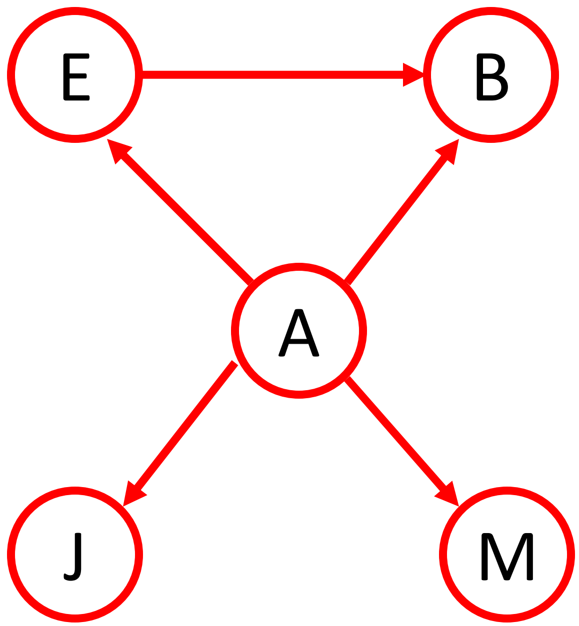
Uniqueness of Minimal I-maps

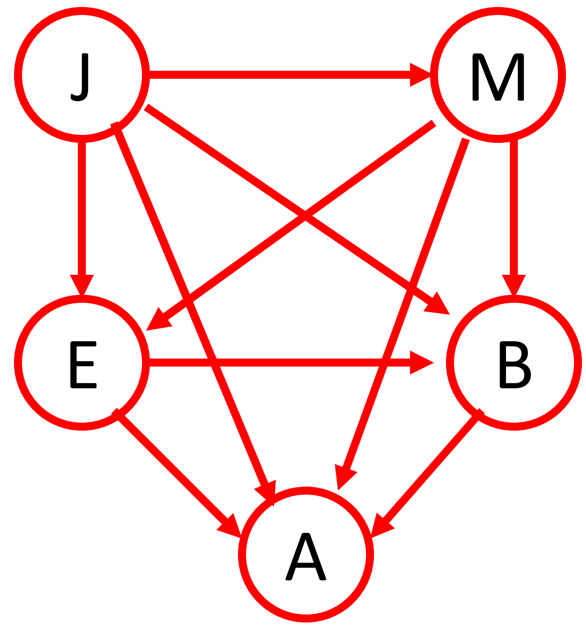
- Is the minimal I-Map unique?

$I(P) = I(G_1)$

G_1







Perfect maps

- Minimal I-maps are easy to find, but can contain many unnecessary dependencies.

- A BN structure G is called **P-map** (perfect map) for distribution P if **I(G) = I(P)**
- Does every distribution P have a P-map?

Existence of perfect maps

$X, Y \sim \mathrm{Ber}(0.5)$

$Z = X \text{ XOR } Y$

$X \perp Y, \quad Y \perp Z, \quad Z \perp X$

$\neg(X \perp Y \; Z)$

$$\begin{tikzcd} X \arrow[dr] & & Y \arrow[dl] \\ & Z & \end{tikzcd}$$

I-map

$\cancel{Y \perp Z}$

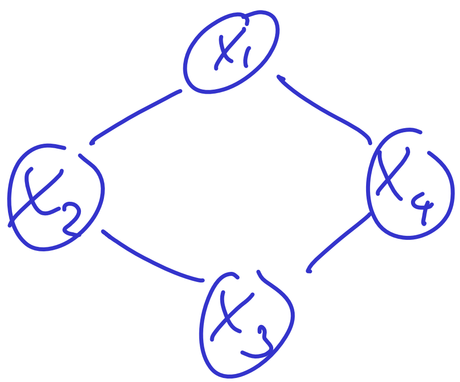
Has no BN P-map

$$\begin{tikzcd} X \arrow[r] & Y \\ Z \arrow[ur] & \end{tikzcd}$$

Existence of perfect maps

$X_1, \ldots X_4$ $\qquad X_1 \perp X_3 \mid X_2\, X_4$

$X_2 \perp X_4 \mid X_1, X_3$



← Undirected GM is a P-map

but NO BN is P-map

Uniqueness of perfect maps

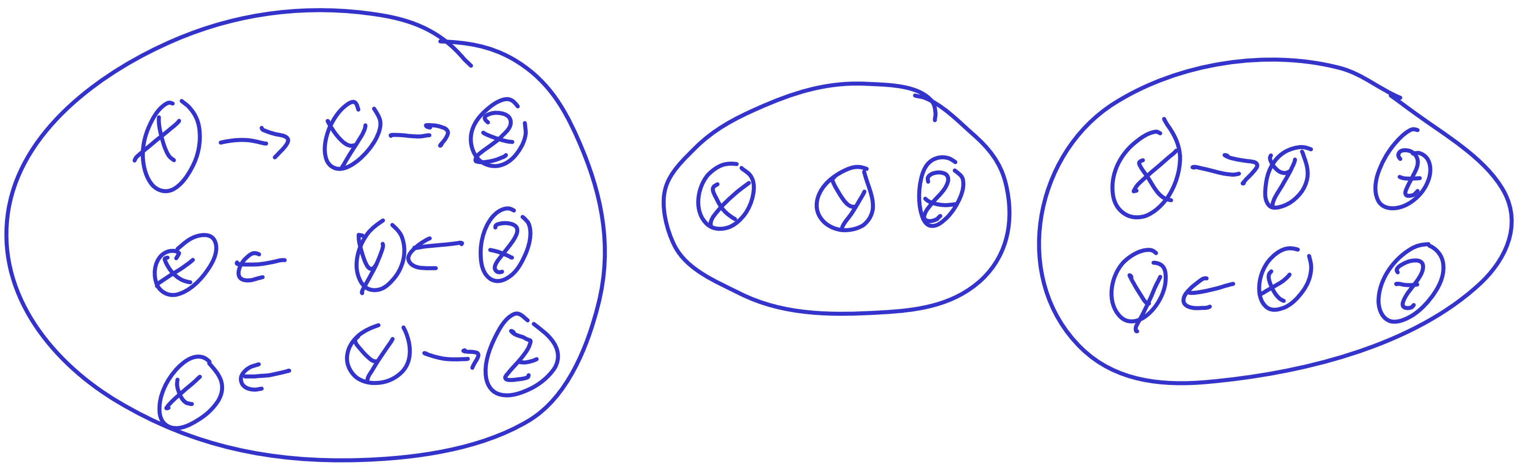
$X \rightarrow Y$ $\quad G_1$

$X \leftarrow Y$ $\quad G_2$

$I(G_1) = I(G_2)$

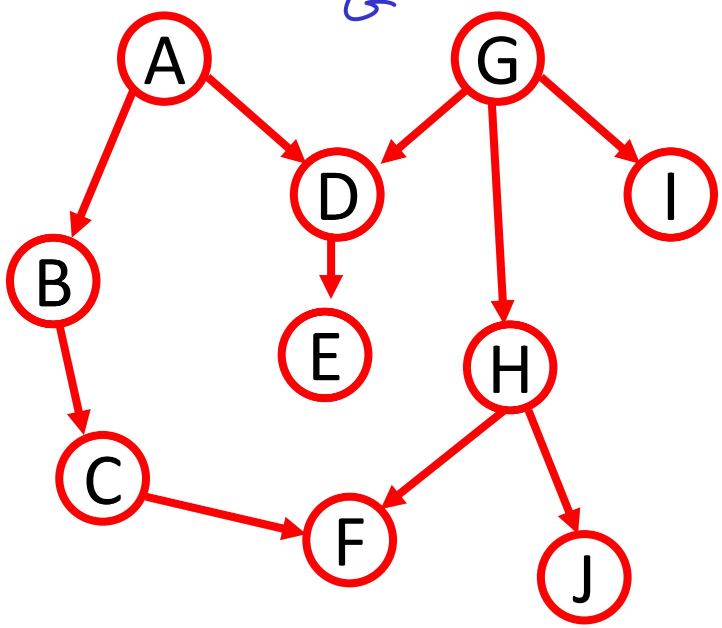
I-Equivalence

- Two graphs G, G’ are called I-equivalent if I(G) = I(G’)
- I-equivalence partitions graphs into equivalence classes

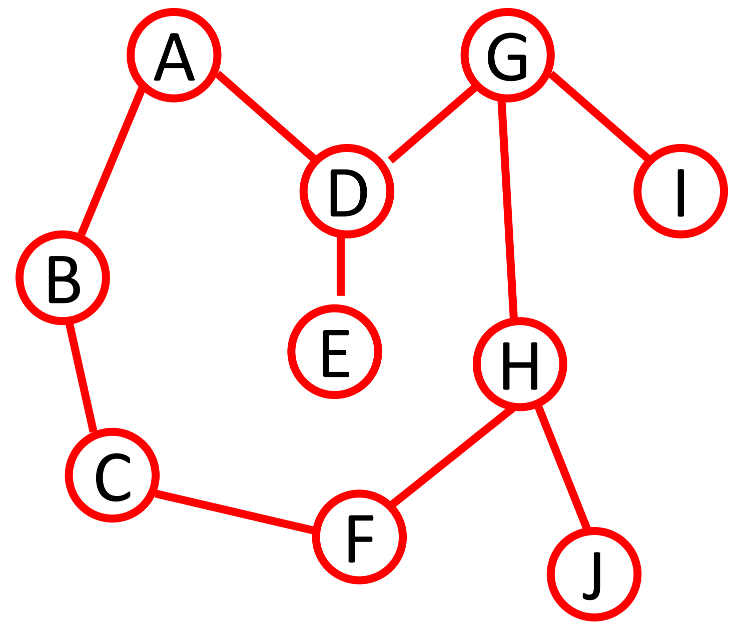


Skeletons of BNs

G

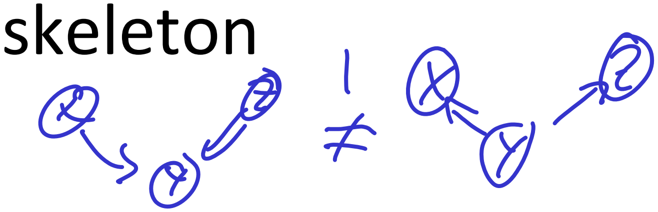


Skeleton of G



- I-equivalent BNs must have same skeleton

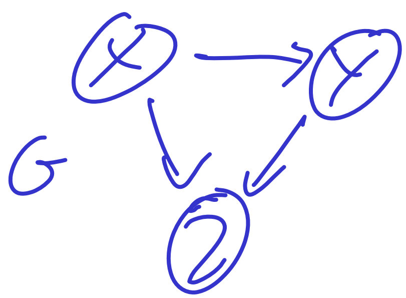
Same skeleton $\not\Rightarrow$ I-equiv.

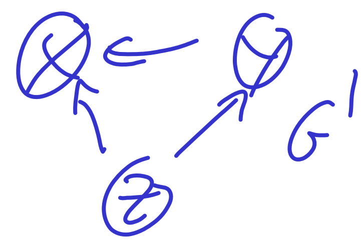


Importance of V-structures

- **Theorem**: If G, G' have same skeleton and same V-structure, then I(G) = I(G')

- Does the converse hold?



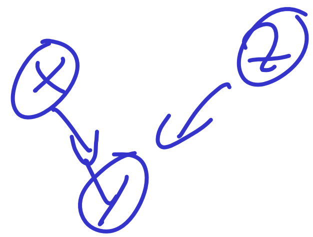


$I(G) = \{\}$ $=$ $I(G') = \{\}$

Same skeleton, NOT same V-structures

Immoralities and I-equivalence

- A V-structure X → Y ← Z is called **immoral** if there is no edge between X and Z ("unmarried parents")



immorality

- **Theorem**: I(G) = I(G') ⇔ G and G' have the same skeleton and the same immoralities.

Tasks

- Subscribe to Mailing list
  https://utils.its.caltech.edu/mailman/listinfo/cs155
- Read Koller & Friedman Chapter 3.3-3.6
- Form groups and think about class projects. If you have difficulty finding a group, email Pete Trautman
- Homework 1 out tonight, due in 2 weeks. Start early!