Reading:
Chapter 2 of Koller&Friedman

BN Semantics 2 –
The revenge of d-separation

Graphical Models – 10708
Carlos Guestrin
Carnegie Mellon University

September 19th, 2005

Announcements

- Homework 1:
  - Out already
  - Due October 3rd – **beginning of class!**
  - It’s hard – start early, ask questions

I-map: $I_\ell(G) \subseteq I(P)$

The BN Representation Theorem

If conditional independencies in BN are subset of conditional independencies in *P*

G is an I map of P

Obtain

P factorizes accord. to G

Joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

P fact. ac. G

If joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

G is I-map P

Then conditional independencies in BN are subset of conditional independencies in *P*

Independencies encoded in BN

- We said: All you need is the local Markov assumption
  - $(X_i \perp NonDescendants_{Xi} \mid \mathbf{Pa}_{Xi})$
- But then we talked about other (in)dependencies
  - e.g., explaining away

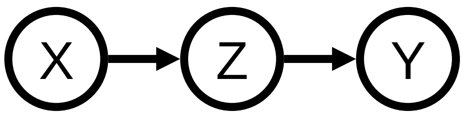
- What are the independencies encoded by a BN?
  - Only assumption is local Markov
  - But many others can be derived using the algebra of conditional independencies!!!

Understanding independencies in BNs – BNs with 3 nodes

obs X

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents

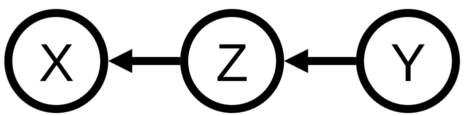
Indirect causal effect:



$(X \perp Y \mid Z)$

not $(X \perp Y)$

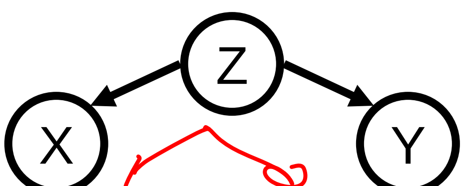
Indirect evidential effect:



$(X \perp Y \mid Z)$

not $(X \perp Y)$

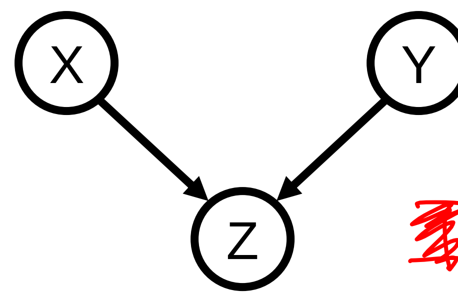
Common cause:



$(X \perp Y \mid Z)$

not $(X \perp Y)$

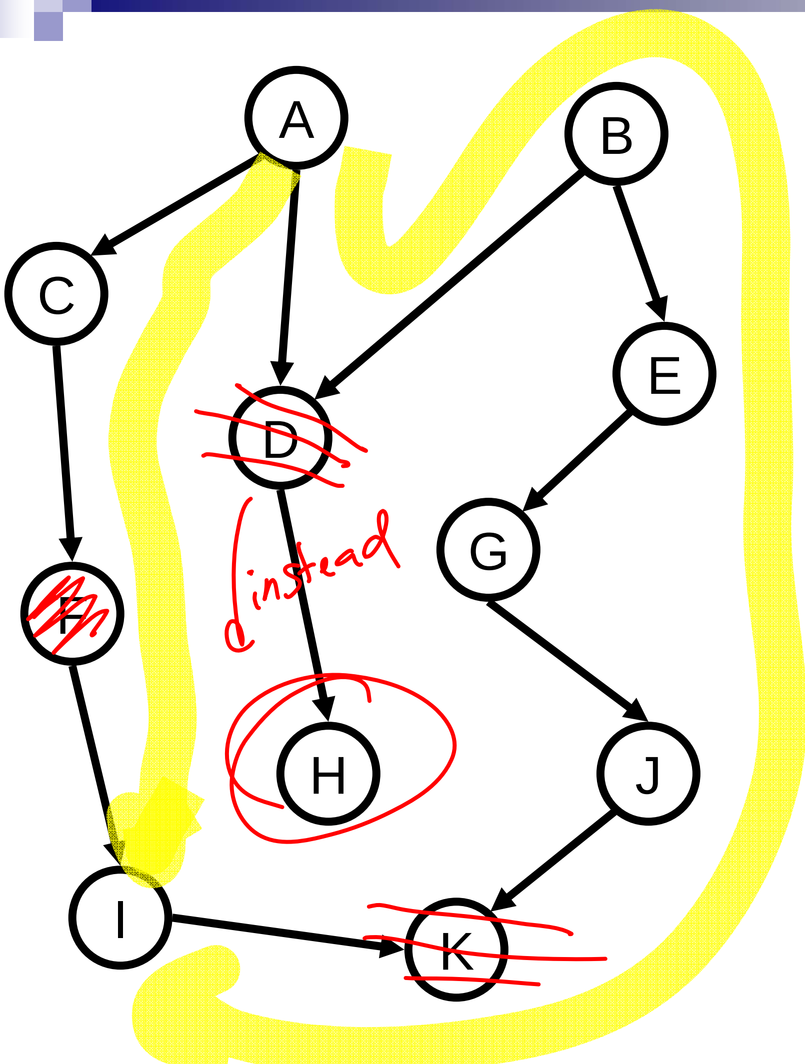
Common effect:



$(X \perp Y)$

not $(X \perp Y \mid Z)$

Understanding independencies in BNs – Some examples



$(H \perp A \mid D)$

$(A \perp B)$

not $(A \perp B \mid D)$

not $(A \perp B \mid H)$

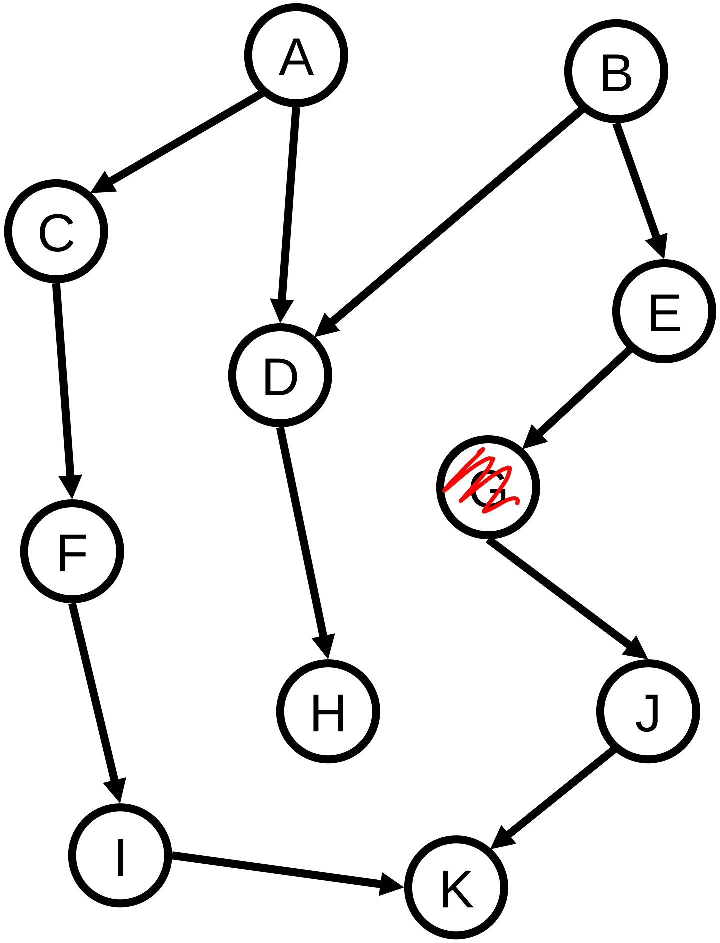
not $(A \perp B \mid K)$

1: $(A \perp I \mid F)$

2 not $(A \perp I \mid F, D, K)$

3 not $(A \perp I \mid F, H, K)$

Understanding independencies in BNs – Some more examples



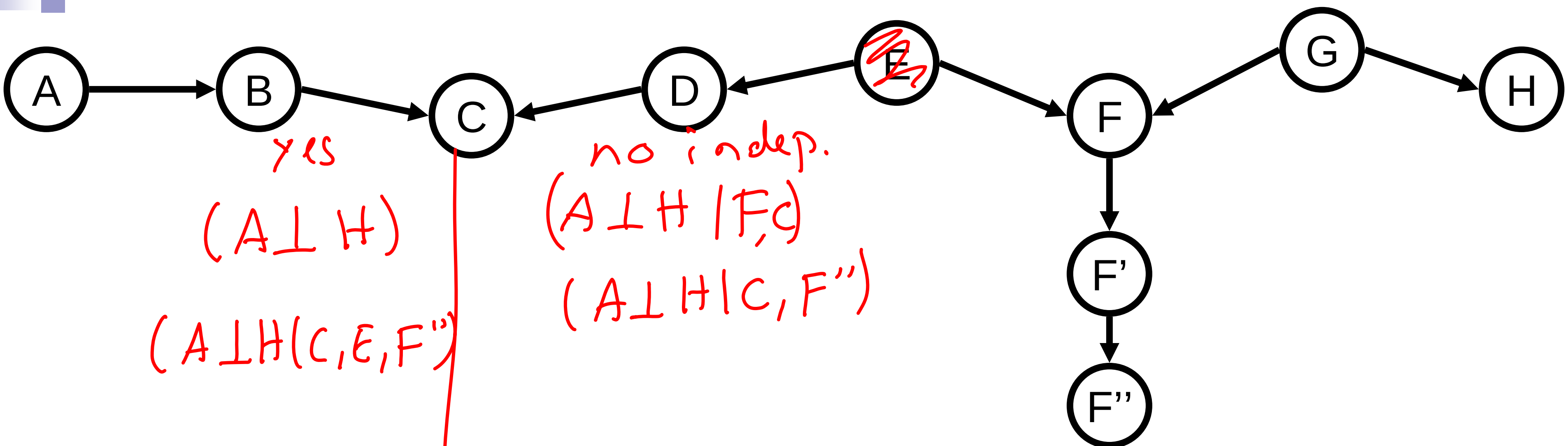
$(A \perp B)$

not $(A \perp B \mid H)$

not $(A \perp B \mid K)$

$(A \perp B \mid K, G)$

An active trail – Example



yes

$(A \perp H)$

$(A \perp H \mid C, E, F'')$

no indep.

$(A \perp H \mid F, C)$

$(A \perp H \mid C, F'')$

When are A and H independent?

Active trails formalized

- A ~~path~~ trail $X_1 - X_2 - \cdots - X_k$ is an **active trail** when variables $\boldsymbol{O} \subseteq \{X_1, \ldots, X_n\}$ are observed if for each consecutive triplet in the trail:
  - $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$, and X_i is **not observed** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, and X_i is **not observed** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, and X_i is **not observed** ($X_i \notin \boldsymbol{O}$)

V- structure

  - $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, and X_i **is observed** ($X_i \in \boldsymbol{O}$), or **one of its descendents**

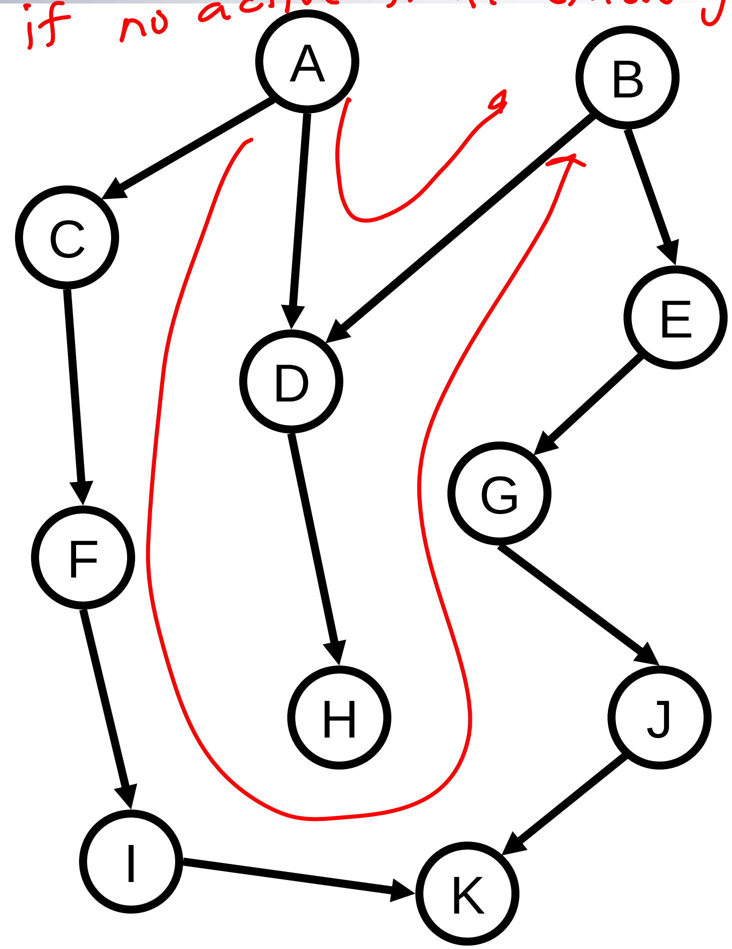
Active trails and independence?

X_i and X_j are d-sep given Z, if no active trail exists given Z

- **Theorem**: Variables $\mathbf{X_i}$ **and** $\mathbf{X_j}$ **are independent given** $\boldsymbol{Z} \subseteq \{X_1,\ldots,X_n\}$ if the is **no active trail** between X_i and X_j when variables $\boldsymbol{Z} \subseteq \{X_1,\ldots,X_n\}$ are observed:
  - i.e., $(\mathbf{X_i} \perp \mathbf{X_j} \mid \mathbf{Z}) \subseteq I(P)$

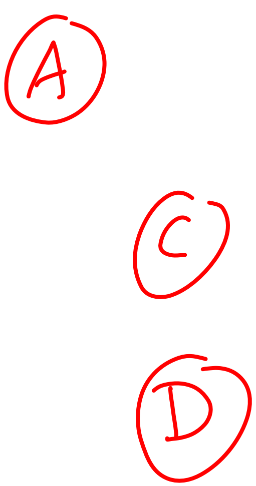
$(A \perp B)$

not $(A \perp B \mid K)$



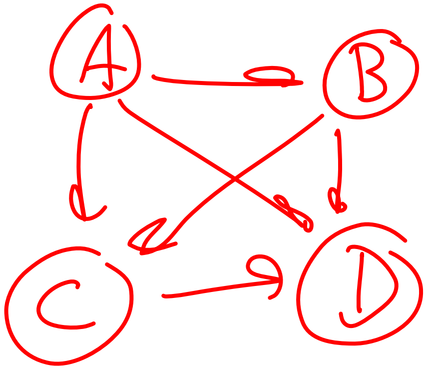
Two interesting (trivial) special cases

Edgeless Graph

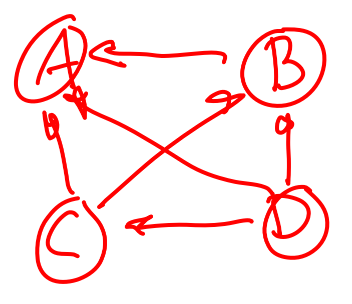


all vars (and subsets) indep.

Complete Graph



no indep.



More generally: Soundness of d-separation

local Markov assump. $I_\ell(G)$

- Given BN structure *G*
- Set of independence assertions obtained by d-separation:
  - $I(G) = \{(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}) : \text{d-sep}_G(\mathbf{X};\mathbf{Y}|\mathbf{Z})\}$
- **Theorem: Soundness of d-separation**
  - If *P* factorizes over *G* then $I(G) \subseteq I(P)$
- **Interpretation:** d-separation only captures true independencies
- Proof discussed when we talk about undirected models

Existence of dependency when not d-separated

not $\text{d-sep}_G(A \perp B \mid H)$

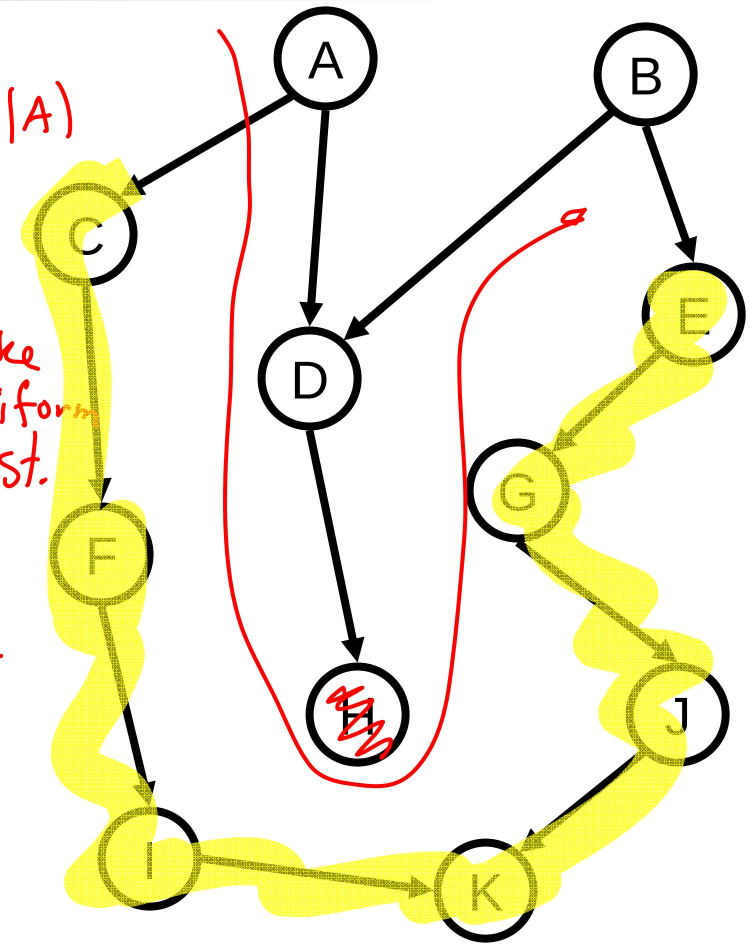
- **Theorem:** If X and Y are not d-separated given **Z**, then X and Y are dependent given **Z** under some *P* that factorizes over *G*
- **Proof sketch**:
  - Choose an active trail between X and Y given **Z**
  - Make this trail dependent
  - Make all else uniform (independent) to avoid "canceling" out influence

P(C|A)

	A	¬A
C	θ	θ
¬C	1-θ	1-θ

P(C|A)

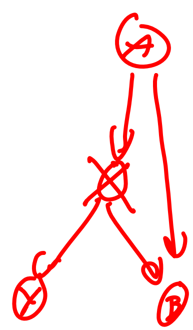
make uniform dist.



Add edges doesn't hurt

$$I_\ell(G) \subseteq I(P) \longrightarrow P \text{ factorizes acc. to } G$$

Start with G where $I_\ell(G) \subseteq I(P)$



in G $(B \perp A \mid X)$

in G $(B \perp Y \mid X)$

---

in G' $(B \perp Y \mid A, X)$

what if add edge to G call it G'

$$I_\ell(G') \subseteq I(G) \subseteq I(P)$$

More generally: Completeness of d-separation

$A \to B$ $\quad P(A)$

$P(B|A) = P(B) \; \forall A$

- **Theorem: Completeness of d-separation**
  - For "almost all" distributions that *P* factorize over to *G*, we have that I(*G*) = I(*P*)
  - *"almost all" distributions*: except for a set of measure zero of parameterizations of the CPTs (assuming no finite set of parameterizations has positive measure)
- **Proof sketch:** $A \to B$

$P(A=a) = \theta_a$

$P(A=\neg a) = 1-\theta_a$

$P(B=b \mid A=a) = \theta_{b|a}$

$P(B=b \mid A=\neg a) = \theta_{b|\neg a}$

if $(A \perp B)$

$P(A,B) = P(A) \cdot P(B|A) = P(A)P(B) \; \forall A,B$

$\theta_a \cdot \theta_{b|a} = \theta_a [\theta_a \theta_{b|a} + (1-\theta_a)\theta_{b|\neg a}]$

$\forall a, b$

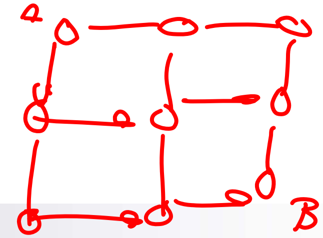
$\vdots$

only if $\theta_{b|a} = \theta_{b|\neg a}$ are $(A \perp B)$

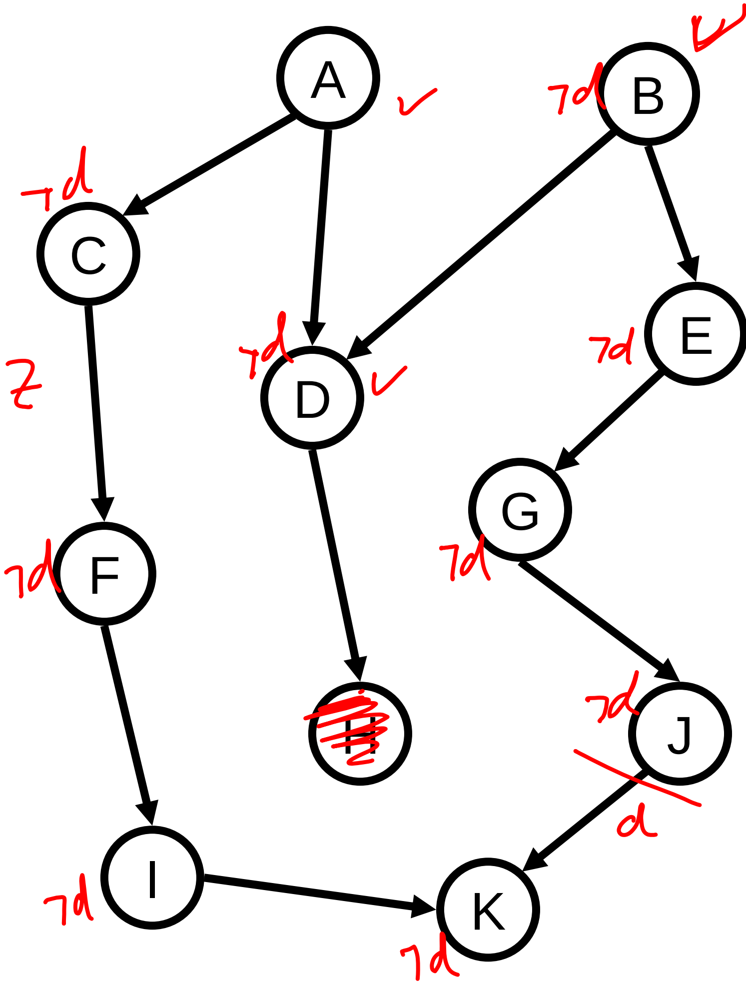
Interpretation of completeness

- **Theorem: Completeness of d-separation**
  - For "almost all" distributions that P factorize over to G, we have that $I(G) = I(P)$
- BN graph is usually sufficient to capture all independence properties of the distribution!!!!
- But only for complete independence:
  - $P \models (\mathbf{X}=\mathbf{x} \perp \mathbf{Y}=\mathbf{y} \mid \mathbf{Z}=\mathbf{z}),\ \forall\, \mathbf{x} \in Val(\mathbf{X}),\ \mathbf{y} \in Val(\mathbf{Y}),\ \mathbf{z} \in Val(\mathbf{Z})$
- Often we have context-specific independence (CSI)
  - $\exists\, \mathbf{x} \in Val(\mathbf{X}),\ \mathbf{y} \in Val(\mathbf{Y}),\ \mathbf{z} \in Val(\mathbf{Z}): P \models (\mathbf{X}=\mathbf{x} \perp \mathbf{Y}=\mathbf{y} \mid \mathbf{Z}=\mathbf{z})$
  - Many factors may affect your grade
  - But if you are a frequentist, all other factors are irrelevant ☺

Algorithm for d-separation



- How do I check if X and Y are d-separated given **Z**
  - There can be exponentially-many trails between X and Y
- Two-pass linear time algorithm finds all d-separations for X given Z
- 1. Upward pass
  - Mark ~~descendants~~ ancestor of **Z**
- 2. Breadth-first traversal from X
  - Stop traversal at a node if trail is "blocked"
  - (Some tricky details apply – see reading)



Building BNs from independence properties

- From d-separation we learned:
  - Start from local Markov assumptions, obtain all independence assumptions encoded by graph
  - For most *P*'s that factorize over *G*, I(*G*) = I(*P*)
  - All of this discussion was for a given *G* that is an I-map for *P*

- Now, give me a *P*, how can I get a *G*?
  - i.e., give me the independence assertions ~~assumptions~~ entailed by *P*
  - Many *G* are "equivalent", how do I represent this?
  - Most of this discussion is not about practical algorithms, but useful concepts that will be used by practical algorithms

Minimal I-maps

- One option:
  - *G* is an I-map for *P*
  - *G* is as simple as possible

- *G* is a **minimal I-map** for *P* if deleting any edges from *G* makes it no longer an I-map

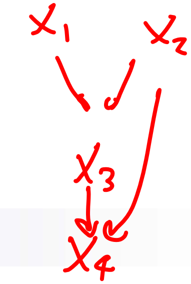
given: $(A \perp B)$

$(A \perp B \mid D)$

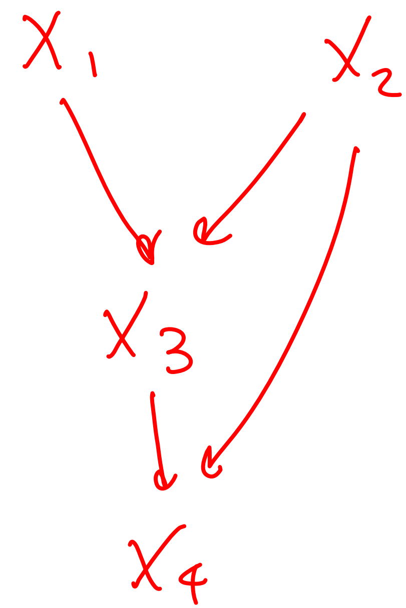
want:

$$\begin{tikzcd} A \arrow[d] & B \\ D \arrow[d] & \\ C & \end{tikzcd}$$

Obtaining a minimal I-map



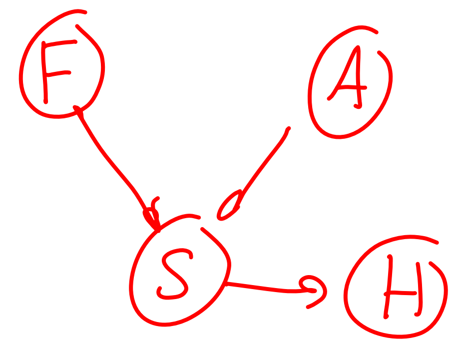
- Given a set of variables and conditional independence assertions ~~assumptions~~
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\ldots,X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1,\ldots,X_{i-1}\}$, given parents $\mathbf{Pa}_{Xi}$
  - Define/learn CPT – $P(X_i| \mathbf{Pa}_{Xi})$



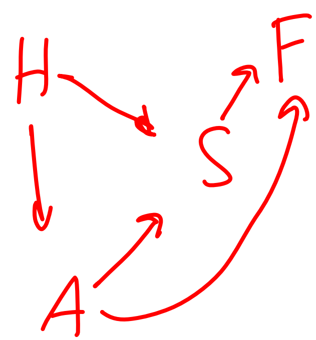
Minimal I-map not unique (or minimal)

- Given a set of variables and conditional independence assumptions
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n
  - Add X_i to the network
  - Define parents of X_i, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\ldots,X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1,\ldots,X_{i-1}\}$, given parents $\mathbf{Pa}_{Xi}$
  - Define/learn CPT – $P(X_i|\ \mathbf{Pa}_{Xi})$

Flu, Allergy, SinusInfection, Headache



order: H A S F



Perfect maps (P-maps)

- I-maps are not unique and often not simple enough

- Define "simplest" G that is I-map for P
  - A BN structure G is a **perfect map** for a distribution P if $I(P) = I(G)$

- Our goal:
  - Find a perfect map!
  - Must address equivalent BNs

Inexistence of P-maps 1

- XOR (this is a hint for the homework)

A, B, C binary : $C = A \text{ xor } B$

A, B uniform (50/50)

$(A \perp B)$

$(A \perp C)$

$(B \perp C)$

not ind:

$(A \perp B \mid C)$

$(A \perp C \mid B)$

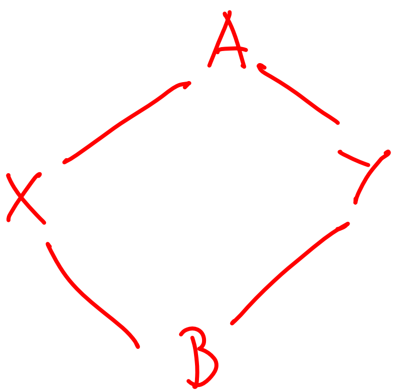
$(B \perp C \mid A)$

I maps	P-maps
$\begin{tikzcd} A \arrow[dr] & & B \arrow[dl] \\ & C & \end{tikzcd}$ $\quad$ $\begin{tikzcd} B \arrow[dr] & & C \arrow[dl] \\ & A & \end{tikzcd}$	NO!

Inexistence of P-maps 2

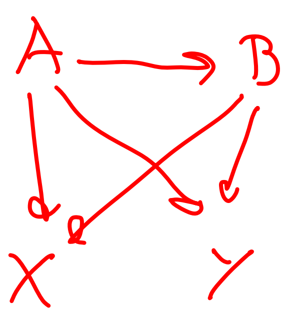
- (Slightly un-PC) swinging couples example

men A, B   women X, Y



$(A \perp B \mid X, Y)$

$(X \perp Y \mid A, B)$

I-map	P-map
	No!

Obtaining a P-map

- Given the independence assertions that are true for *P*

- Assume that there exists a perfect map G^*
  - Want to find G^*

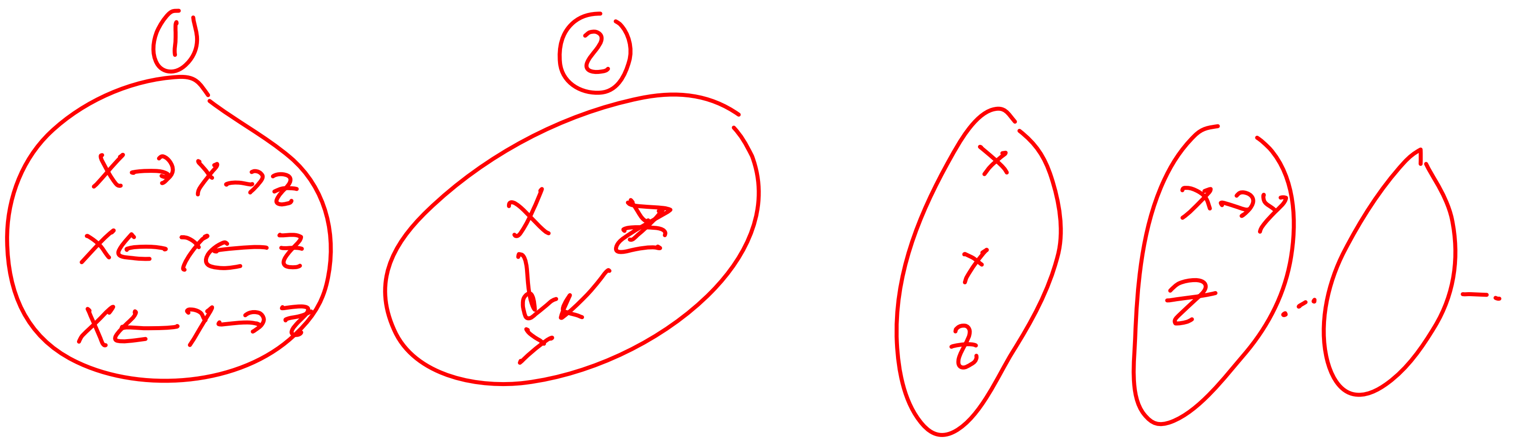
- Many structures may encode same independencies as G^*, when are we done?
  - Find all equivalent structures simultaneously!

I-Equivalence

$X \rightarrow Y \rightarrow Z$

$X \leftarrow Y \rightarrow Z$

- Two graphs G_1 and G_2 are **I-equivalent** if $I(G_1) = I(G_2)$
- **Equivalence class** of BN structures
  - Mutually-exclusive and exhaustive partition of graphs



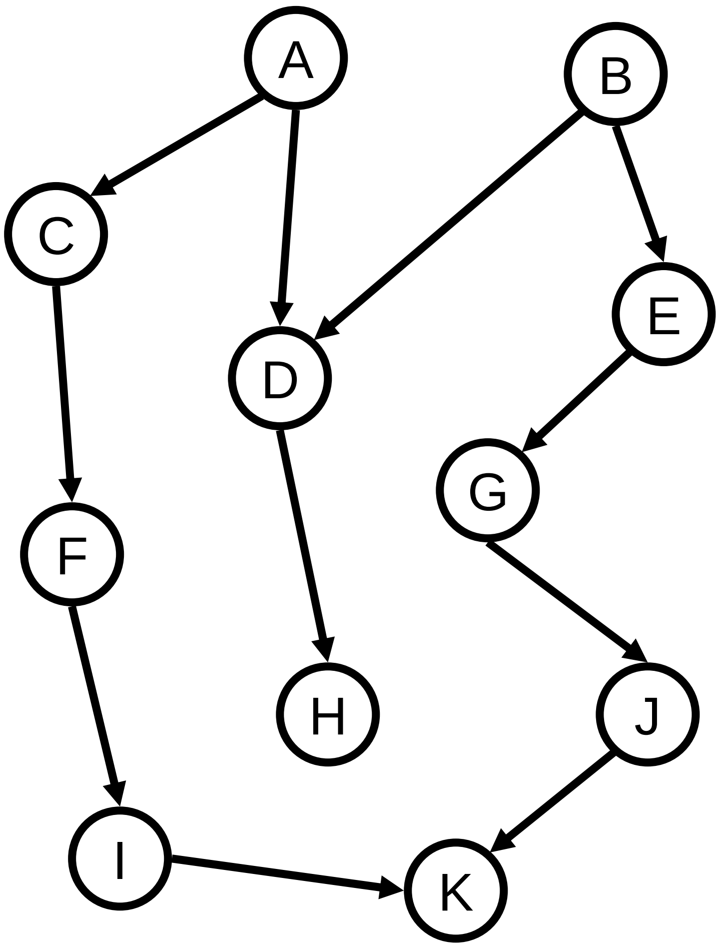
- How do we characterize these equivalence classes?

Skeleton of a BN

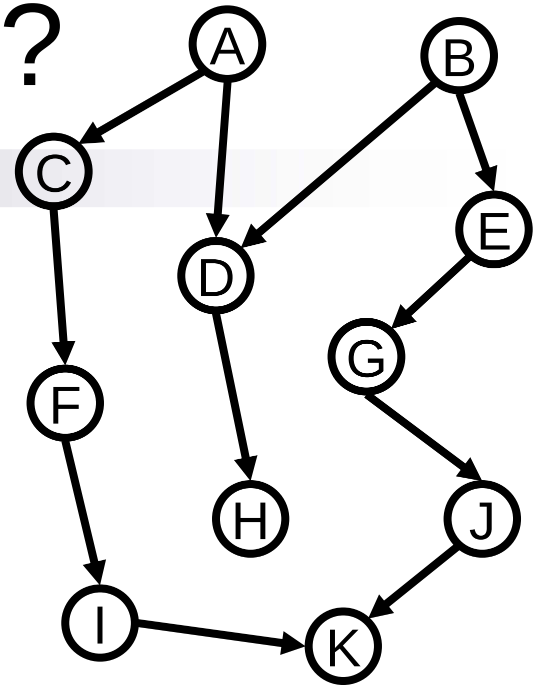
- **Skeleton** of a BN structure *G* is an **undirected graph** over the same variables that has an edge X–Y for every X→Y or Y→X in *G*

- (Little) **Lemma:** Two I-equivalent BN structures must have the same skeleton

counter example



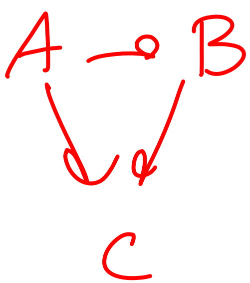
What about V-structures?

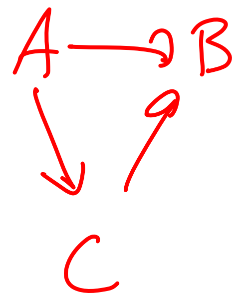


- **V-structures are key property of BN structure**
- **Theorem:** If G_1 and G_2 have the same skeleton and V-structures, then G_1 and G_2 are I-equivalent

Same V-structures not necessary

- **Theorem:** If G_1 and G_2 have the same skeleton and V-structures, then G_1 and G_2 are I-equivalent
- Though sufficient, same V-structures not necessary





V-structure $A \rightarrow C \leftarrow B$

$A \rightarrow B \leftarrow C$

diff. V-structures

same indep.

!

V-structures sufficient not necessary

Immoralities & I-Equivalence

- Key concept not V-structures, but “immoralities” (unmarried parents ☺)
  - $X \rightarrow Z \leftarrow Y$, with no arrow between X and Y
  - Important pattern: X and Y independent given their parents, but not given Z
  - (If edge exists between X and Y, we have *covered* the V-structure)
- **Theorem:** G_1 and G_2 have the same skeleton and immoralities if and only if G_1 and G_2 are I-equivalent

Obtaining a P-map

- Given the independence assertions that are true for *P*
  - Obtain skeleton
  - Obtain immoralities
- From skeleton and immoralities, obtain every (and any) BN structure from the equivalence class

Identifying the skeleton 1

- When is there an edge between X and Y?

- When is there no edge between X and Y?

Identifying the skeleton 2

- Assume d is max number of parents (d could be n)

- For each X_i and X_j
  - $E_{ij} \leftarrow$ true
  - For each $\mathbf{U} \subseteq \mathbf{X} - \{X_i, X_j\}$, $|\mathbf{U}| \leq 2d$
    - Is $(X_i \perp X_j \mid \mathbf{U})$?
      - $E_{ij} \leftarrow$ true
  - If E_{ij} is true
    - Add edge X – Y to skeleton

Identifying immoralities

- Consider X – Z – Y in skeleton, when should it be an immorality?

- Must be $X \rightarrow Z \leftarrow Y$ (immorality):
  - When X and Y are **never independent** given **U,** if $Z \in$ **U**

- Must **not** be $X \rightarrow Z \leftarrow Y$ (not immorality):
  - When there exists **U** with $Z \in$ **U**, such that X and Y are **independent** given **U**

From immoralities and skeleton to BN structures

- Representing BN equivalence class as a **partially-directed acyclic graph** (PDAG)
- Immoralities force direction on other BN edges
- Full (polynomial-time) procedure described in reading

What you need to know

- Definition of a BN
- Local Markov assumption
- The representation theorem: *G* is an I-map for *P* if and only if *P* factorizes according to *G*
- d-separation – sound and complete procedure for finding independencies
  - (almost) all independencies can be read directly from graph without looking at CPTs
- Minimal I-map
  - every *P* has one, but usually many
- Perfect map
  - better choice for BN structure
  - not every *P* has one
  - can find one (if it exists) by considering I-equivalence
  - Two structures are I-equivalent if they have same skeleton and immoralities