

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

The BN Representation Theorem

If conditional independencies in BN are subset of conditional independencies in \mathcal{P}

G is an I-map of P

Obtain

P factorizes accord.
Joint probability to G distribution:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Pa}_{X_i})$$

P fact. ac. G
If joint probability distribution:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Pa}_{X_i})$$

Obtain

G is I-map P
Then conditional independencies in BN are subset of conditional independencies in \mathcal{P}

Independencies encoded in BN

- We said: All you need is the local Markov assumption
 - $(X_i \perp \text{NonDescendants}_{X_i} \mid \mathbf{Pa}_{X_i})$
- 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!!!

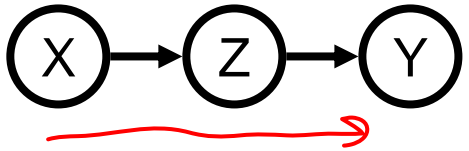
Understanding independencies in BNs

– BNs with 3 nodes

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

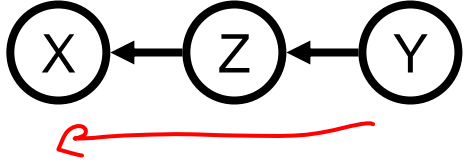
obs X

Indirect causal effect:



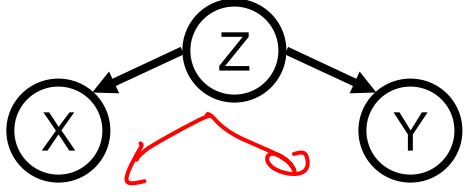
$(X \perp Y | Z)$
not $(X \perp Y)$

Indirect evidential effect:



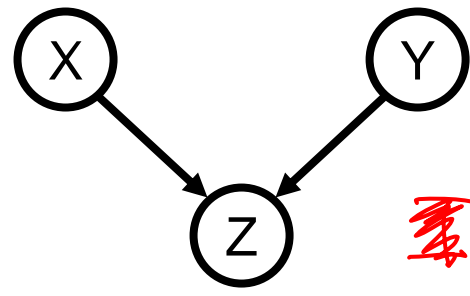
$(X \perp Y | Z)$
not $(X \perp Y)$

Common cause:



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

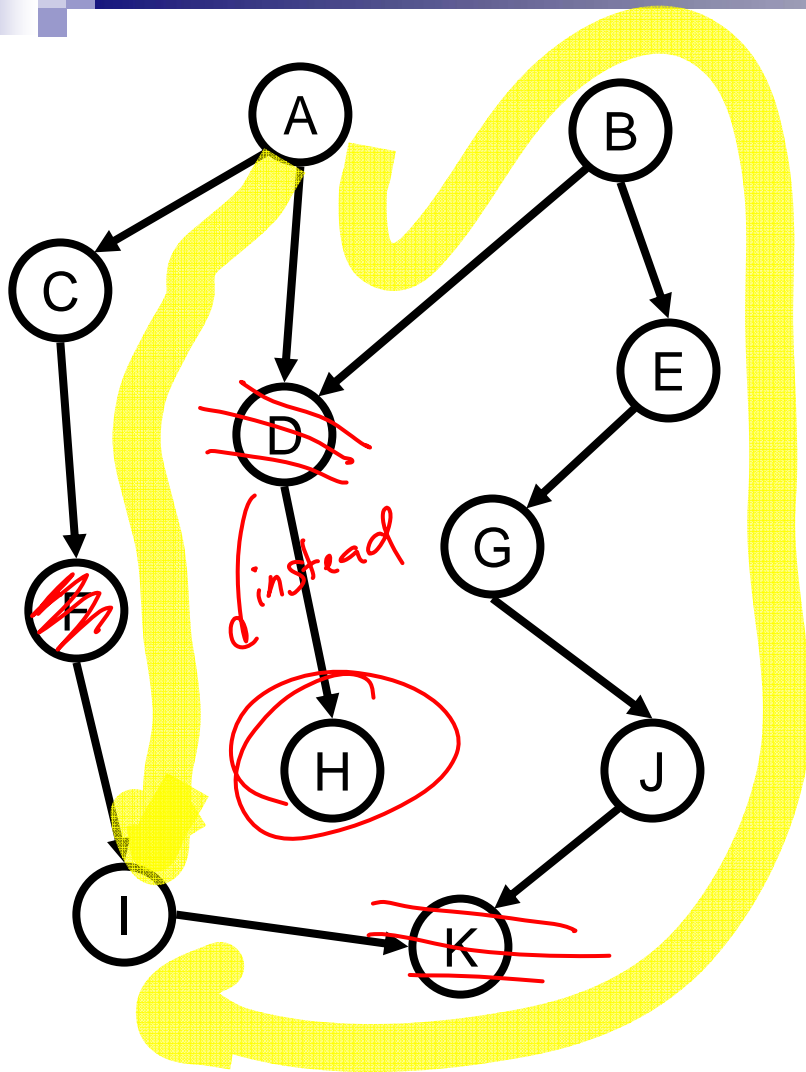
Common effect:



~~$(X \perp Y)$~~
not $(X \perp Y | Z)$

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 $(A \perp B \mid K)$

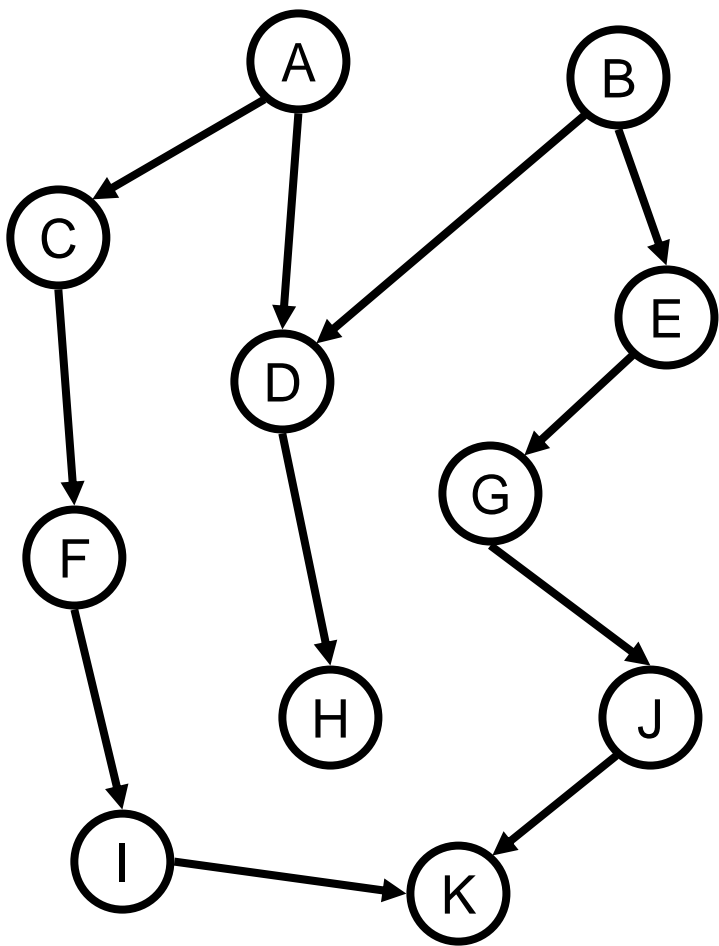
①: $(A \perp I \mid F)$

② not $(A \perp I \mid F, D, K)$

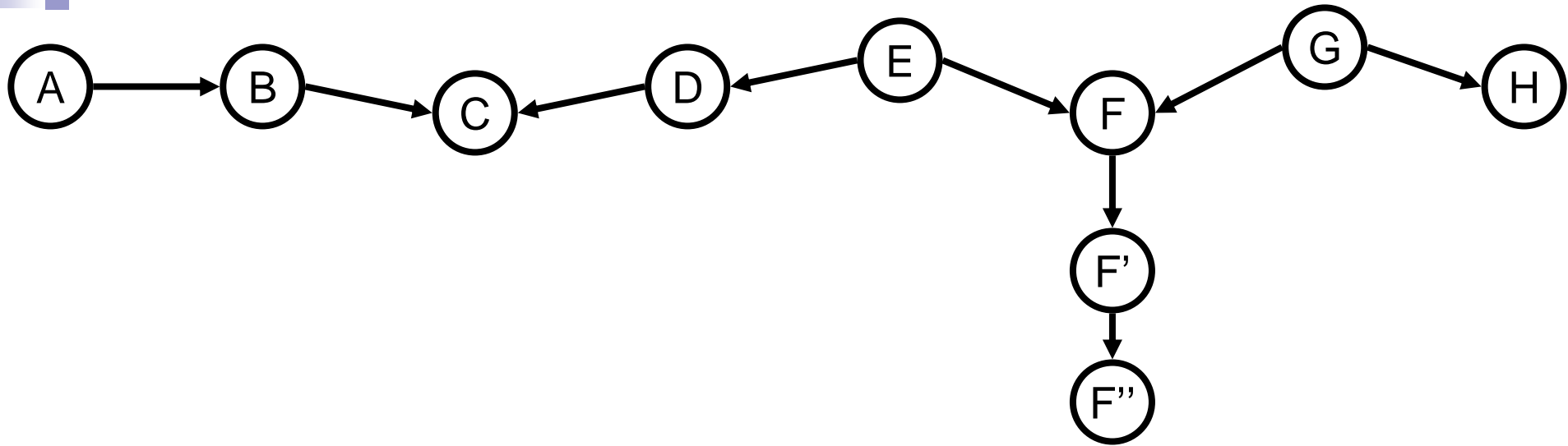
③ not $(A \perp I \mid F, H, K)$

Understanding independencies in BNs

- Some more examples



An active trail – Example



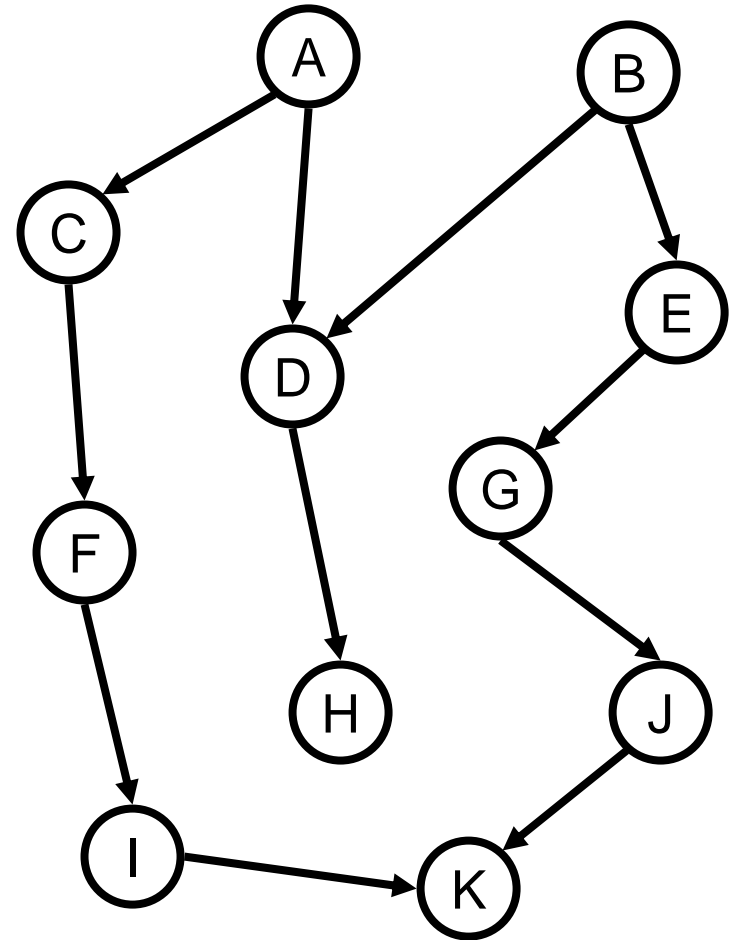
When are A and H independent?

Active trails formalized

- A path $X_1 - X_2 - \dots - X_k$ is an **active trail** when variables $\mathbf{O} \subseteq \{X_1, \dots, X_n\}$ 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** ($X_i \notin \mathbf{O}$)
 - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, and X_i is **not observed** ($X_i \notin \mathbf{O}$)
 - $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, and X_i is **not observed** ($X_i \notin \mathbf{O}$)
 - $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, and X_i is **observed** ($X_i \in \mathbf{O}$), or **one of its descendants**

Active trails and independence?

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



Two interesting (trivial) special cases



Edgeless Graph

Complete Graph

More generally:

Soundness of d-separation

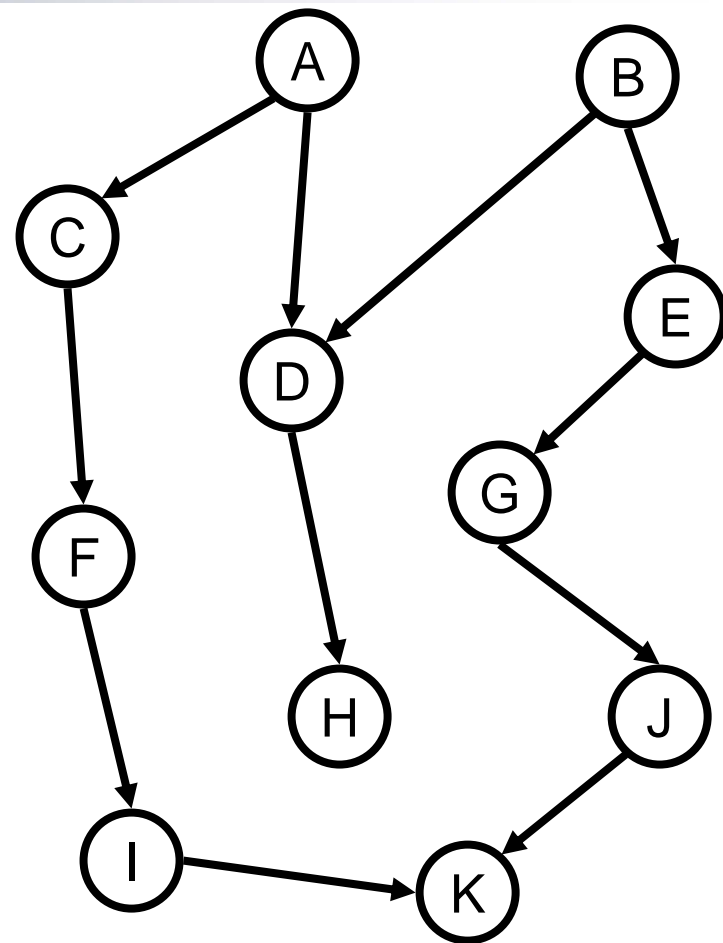
- Given BN structure G
- Set of independence assertions obtained by d-separation:
 - $I(G) = \{(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}) : \text{d-sep}_G(\mathbf{X}; \mathbf{Y} | \mathbf{Z})\}$
- **Theorem: Soundness of d-separation**
 - If P factorizes over G then $I(G) \subseteq I(P)$
- **Interpretation:**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

Existence of dependency when not d-separated

■ **Theorem:** If X and Y are not d-separated given \mathbf{Z} , then X and Y are dependent given \mathbf{Z} under some P that factorizes over G

■ **Proof sketch:**

- Choose an active trail between X and Y given \mathbf{Z}
- Make this trail dependent
- Make all else uniform (independent) to avoid “canceling” out influence



More generally:

Completeness of d-separation

■ 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:

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:

□ $P \models (\mathbf{X}=\mathbf{x} \perp \mathbf{Y}=\mathbf{y} \mid \mathbf{Z}=\mathbf{z}), \forall \mathbf{x} \in \text{Val}(\mathbf{X}), \mathbf{y} \in \text{Val}(\mathbf{Y}), \mathbf{z} \in \text{Val}(\mathbf{Z})$

■ Often we have context-specific independence (CSI)

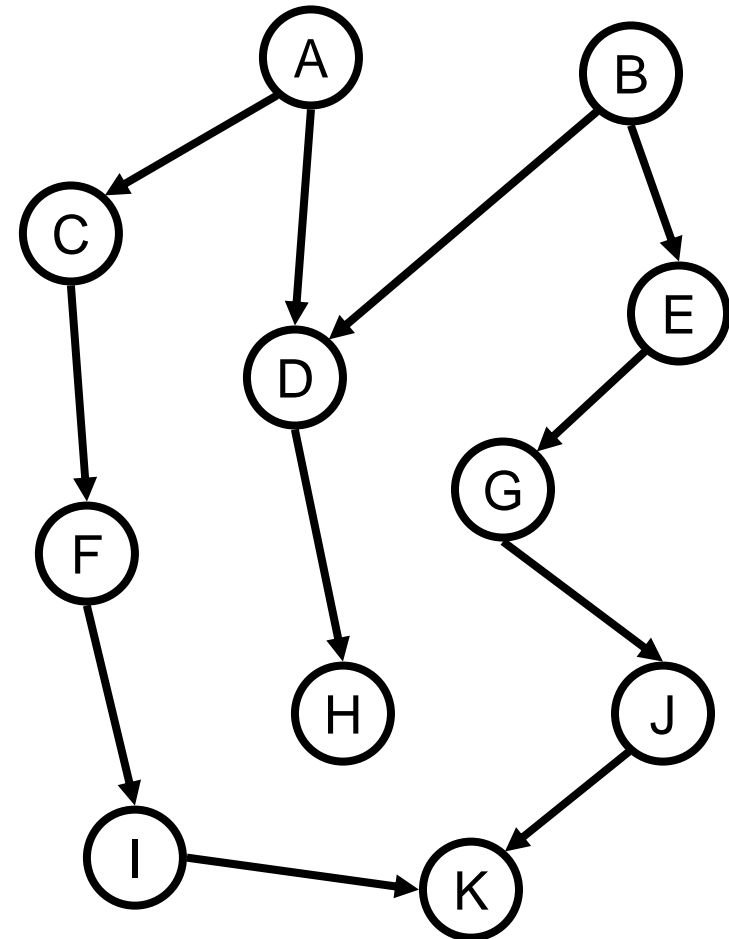
□ $\exists \mathbf{x} \in \text{Val}(\mathbf{X}), \mathbf{y} \in \text{Val}(\mathbf{Y}), \mathbf{z} \in \text{Val}(\mathbf{Z}): P \models (\mathbf{X}=\mathbf{x} \perp \mathbf{Y}=\mathbf{y} \mid \mathbf{Z}=\mathbf{z})$

□ 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
- 1. Upward pass
 - Mark descendants 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 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

Obtaining a minimal I-map

- Given a set of variables and conditional independence assumptions
- Choose an ordering on variables, e.g., X_1, \dots, 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 $\{X_1, \dots, X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1, \dots, X_{i-1}\}$, given parents \mathbf{Pa}_{X_i}
 - Define/learn CPT – $P(X_i | \mathbf{Pa}_{X_i})$

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, \dots, 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 $\{X_1, \dots, X_{i-1}\}$ such that local Markov assumption holds – X_i independent of rest of $\{X_1, \dots, X_{i-1}\}$, given parents \mathbf{Pa}_{X_i}
 - Define/learn CPT – $P(X_i | \mathbf{Pa}_{X_i})$

Flu, Allergy, SinusInfection, Headache

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)

Inexistence of P-maps 2

- (Slightly un-PC) swinging couples example

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!

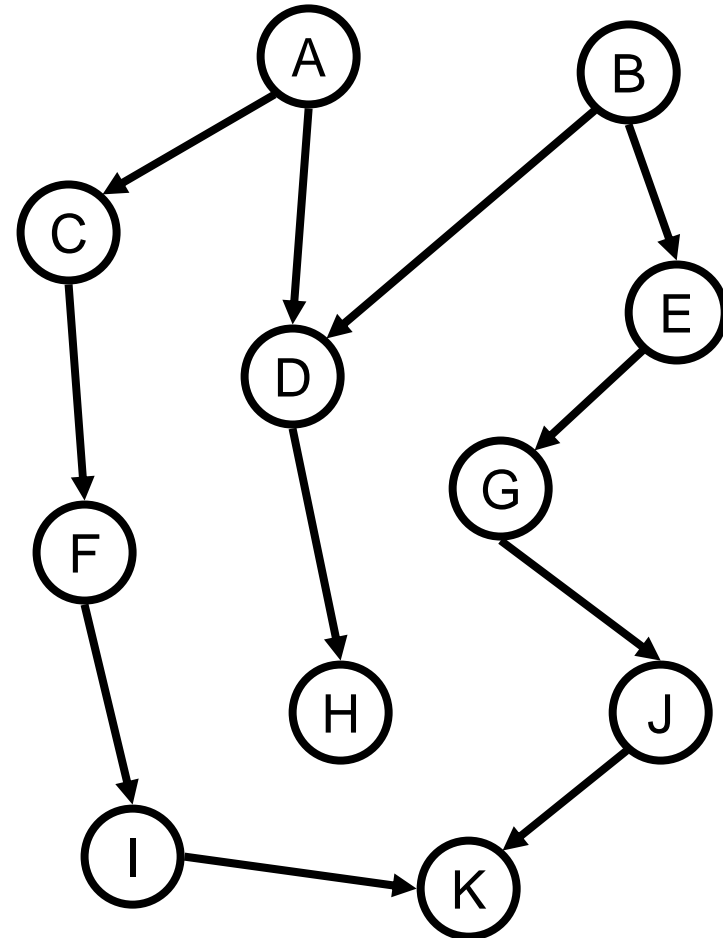
I-Equivalence

- 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

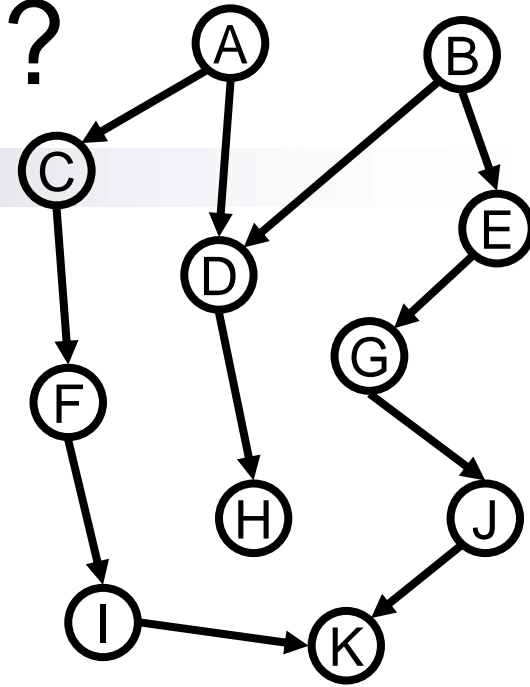
- **Skeleton** of a BN structure G is an **undirected graph** over the same variables that has an edge $X-Y$ for every $X \rightarrow Y$ or $Y \rightarrow X$ in G
- (Little) **Lemma**: Two I-equivalent BN structures must have the same skeleton



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

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 2

- Assume d is max number of parents (d could be n)
- For each X_i and X_j
 - $E_{ij} \leftarrow \text{true}$
 - For each $\mathbf{U} \subseteq \mathbf{X} - \{X_i, X_j\}$, $|\mathbf{U}| \leq 2d$
 - Is $(X_i \perp X_j \mid \mathbf{U})$?
 - $E_{ij} \leftarrow \text{true}$
 - 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 \mathbf{U} , if $Z \in \mathbf{U}$
- Must **not** be $X \rightarrow Z \leftarrow Y$ (not immorality):
 - When there exists \mathbf{U} with $Z \in \mathbf{U}$, such that X and Y are **independent** given \mathbf{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