Readings:
K&F: 3.3, 3.4

# BN Semantics 3 – Now it's personal!


Graphical Models – 10708
Carlos Guestrin
Carnegie Mellon University

September 22nd, 2008



## 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

$A \to B \to C \to D$

$A \perp D \mid B$

$A \to C \leftarrow B$

$A \perp B$

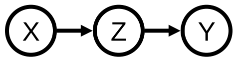
$\neg A \perp B \mid C$

- 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 and only its parents

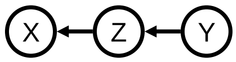
Indirect causal effect:




$Y \perp X \mid Z$

$\neg X \perp Y$

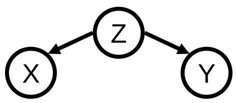
Indirect evidential effect:




$Y \perp X \mid Z$

$\neg X \perp Y$

Common cause:



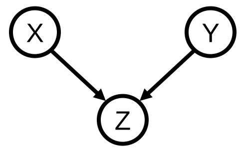

$Y \perp X \mid Z$

$\neg X \perp Y$

all represent same dist.

v-structures

Common effect:

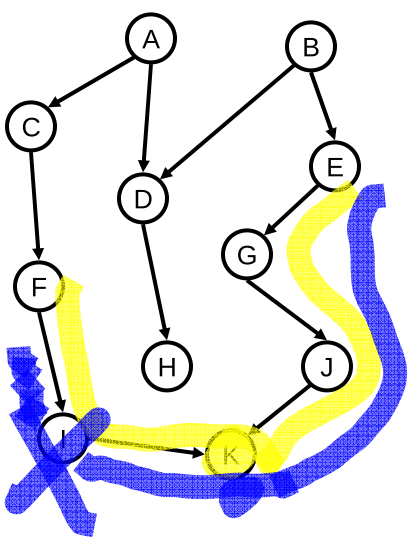



$X \perp Y$

$\neg X \perp Y \mid Z$



# Understanding independencies in BNs – Some examples




$A \perp B$ $\quad F \perp \{B, E, G, J\}$

$I \perp J$ $\quad \{A, C, F, I\} \perp \{B, E, G, J\}$

$B \perp G \mid E$

$B \perp J \mid E$

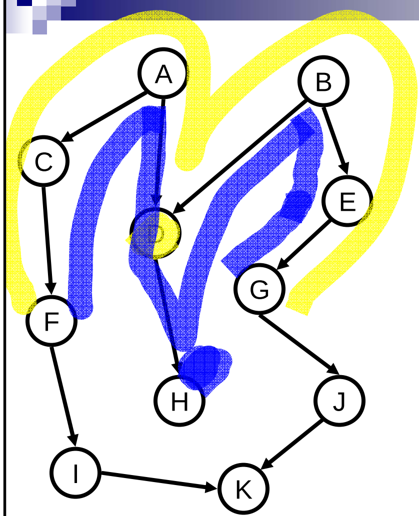
$\neg E \perp F \mid K$

$E \perp F \mid K, I$

difference

$F \perp K \mid I$

# Understanding independencies in BNs – Some more examples




$F \perp G$

$\neg F \perp G \mid D$

$\neg F \perp G \mid H$

$\neg F \perp G \mid H K$

$F \perp G \mid H$ A B C E

A F
$b \searrow \ \swarrow d$
S
$\downarrow$
H

$A \perp F$

$\neg A \perp F \mid S$

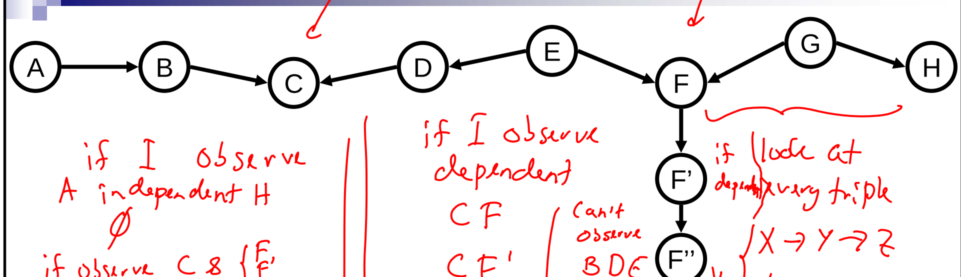
$\neg A \perp F \mid H$



# An active trail – Example

v-structure

v-structure




if I observe
A independent H
$\emptyset$

if observe C & {F, F', F''}
I must observe at least
one of: B D E G

if I observe
dependent
C F
C F'
C F''

can't observe
B D E
G

if look at
dependency every triple
{ X → Y → Z
X ← Y ← Z
X ← Y → Z
Y not observed

**When are A and H independent?**

A → B → C → E
B ↘ D ↗ C

there are
2 trails
in this case

Y or
a descendant
is observed
X ↘ Y ↙ Z

# Active trails formalized

trail is undirected path that never visits a node twice

- A trail $X_1 - X_2 - \cdots - X_k$ is an **active trail** when variables $\boldsymbol{O} \subseteq \{X_1,\dots,X_n\}$ 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** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, 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** ($X_i \notin \boldsymbol{O}$)

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

    v-structure

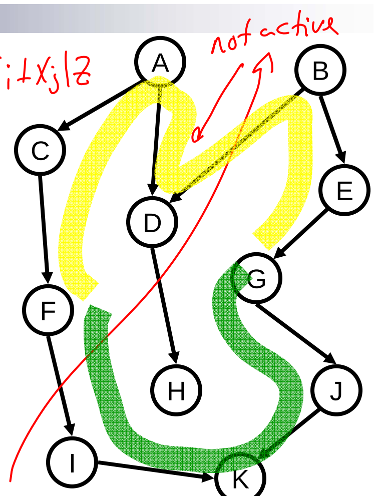


# Active trails and independence?

- **Theorem**: Variables (sets of) $\mathbf{X_i}$ **and** $\mathbf{X_j}$ **are independent given** $\boldsymbol{Z} \subseteq \{X_1,\dots,X_n\}$ if the is **no active trail** between $X_i$ and $X_j$ when variables $\boldsymbol{Z} \subseteq \{X_1,\dots,X_n\}$ are observed

$X_i \perp X_j \mid Z$

not active




we say that $X_i$ and $X_j$ are d-separated given $Z$ (dependency separation)

$F \perp G \rightsquigarrow$ not active

# More generally: Soundness of d-separation

local Markov assumption $I_\ell(G)$
NON Descendants doesn't include direct parents

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

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

not only $I_\ell(G) \subseteq I(P)$

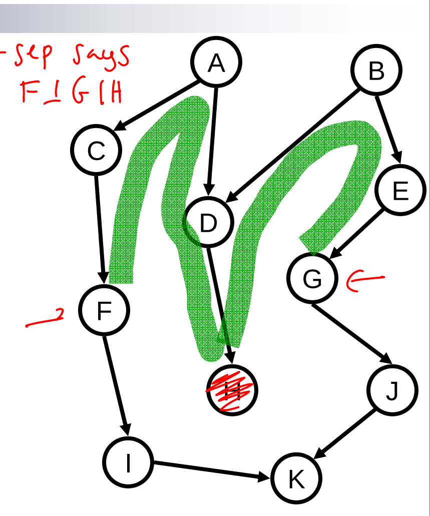
- **Interpretation:** d-separation only captures true independencies
- Proof discussed when we talk about undirected models



# Existence of dependency when not d-separated

- **Theorem:** If X and Y are not d-separated given **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

d-sep says $\neg F \perp G | H$



# More generally: Completeness of d-separation

- **Theorem: Completeness of d-separation**
  - For "almost all" distributions where *P* factorizes 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)
    - Means that if all sets **X** & **Y** that are not d-separated given **Z**, then ¬(**X**⊥**Y**|**Z**)
- **Proof sketch for very simple case:**

$X \rightarrow Y \leftarrow$ $\theta_{Y|X=t}$, $\theta_{Y|X=f}$

$\theta_X = P(X=t)$

d-sep $\neg X \perp Y$

but if $X \perp Y$

$P(Y|X) = P(Y)$

$\underbrace{\theta_{Y|X=t} = \theta_X \cdot \theta_{Y|X=t} + (1-\theta_X)\theta_{Y|X=f}}_{\text{polynomial}}$

finite # of roots

measure is $\emptyset$

OR linear subspace also has $\emptyset$ measure



# Interpretation of completeness

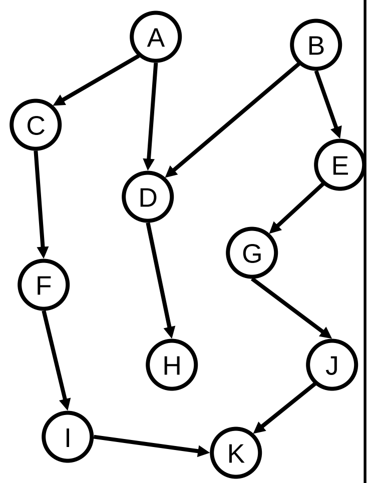
e.g. decision trees

- **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* ➔(**X**=**x**⊥**Y**=**y** | **Z**=**z**), ∀ **x**∈Val(**X**), **y**∈Val(**Y**), **z**∈Val(**Z**)
- Often we have context-specific independence (CSI)
  - ∃ **x**∈Val(**X**), **y**∈Val(**Y**), **z**∈Val(**Z**): *P* ➔(**X**=**x**⊥**Y**=**y** | **Z**=**z**)
  - Many factors may affect your grade
  - But if you are a frequentist, all other factors are irrelevant ☺

EXTREMELY USEFUL FOR ML

# 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)






# What you need to know

- d-separation and independence $I(G) \subseteq I(P)$
  - sound procedure for finding independencies
  - existence of distributions with these independencies
  - (almost) all independencies can be read directly from graph without looking at CPTs $I(G) = I(P)$

# Announcements

- Homework 1:
  - Due next Wednesday – **beginning of class!**
  - It's hard – start early, ask questions
- Audit policy
  - No sitting in, official auditors only, see course website

# 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*

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

- Now, give me a *P*, how can I get a *G*?
  - i.e., give me the independence ~~assumptions~~ assertions 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
    - Practical algs next time

# 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

$\neg A \perp B$
$A \perp B \mid C$

$A \to C \to B$

if remove an edge then no longer I map



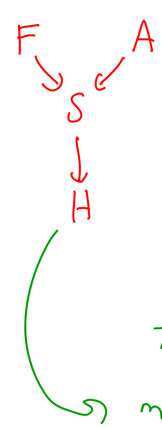
# Obtaining a minimal I-map

- Given a set of variables and conditional independence ~~assumptions~~ assertions
- 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 $\mathbf{Pa}_{X_i}$
  - Define/learn CPT – $P(X_i \mid \mathbf{Pa}_{X_i})$

Flu, Allergy, SinusInfection, Headache

F A S H

$F \perp A$
$F A \perp H \mid S$




if I remove $A \to S$, $\Rightarrow$ $A \perp H$ not true in P

Proof by example, minimal I map

# Minimal I-map not unique (or minimum)

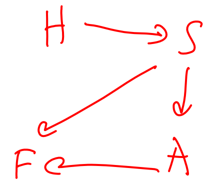
- Given a set of variables and conditional independence assumptions
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n
  - Add $X_i$ to the network
  - Define parents of $X_i$, $\mathbf{Pa}_{X_i}$, in graph as the minimal subset of $\{X_1,\ldots,X_{i-1}\}$ such that local Markov 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

H S A F

$A \perp F$

$A F \perp H | S$




still an I map

more edges

Minimal? Yes!!

doesn't capture $A \perp F$



# 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)

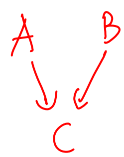
$A = B \text{ XOR } C$

$A \perp B$
$B \perp C$
$C \perp A$

$\neg A \perp B \mid C$
$\neg A \perp C \mid B$
$\neg B \perp C \mid A$

P-MAP?
extra credit

I map :
Minimal
$A \perp B$




not P-map
can't read
$B \perp C$
$A \perp C$



# Inexistence of P-maps 2

- (Slightly un-PC) swinging couples example

1
$\{X_1, Y_1\}$

2
$\{X_2, Y_2\}$

$\neg X_1 \perp Y_1$
$\neg X_1 \perp Y_2$

$X_1 \perp X_2 \mid Y_1, Y_2$

$Y_1 \perp Y_2 \mid X_1, X_2$

P-MAP : undirected Graphical Model


P MAP Bayes Net?
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!