Readings:

K&F: 3.3, 3.4

BN Semantics 2 –

The revenge of d-separation

Graphical Models – 10708

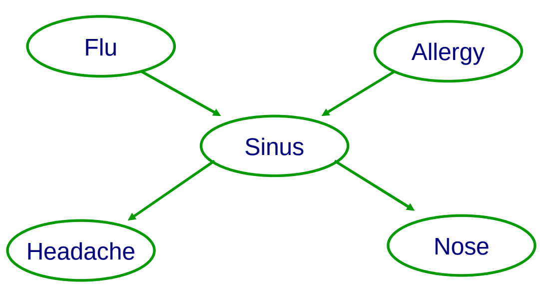
Carlos Guestrin

Carnegie Mellon University

September 20th, 2006

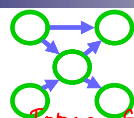
Local Markov assumption & I-maps

- Assumptions
- Local independence assumptions in BN structure G: $I_\ell(G)$
- Truth:
- Independence assertions of *P*: $I(P)$
- BN structure G is an ***I-map*** (independence map) if: for P $I_\ell(G) \subseteq I(P)$



Local Markov Assumption:
A variable X is independent of its non-descendants given its parents
($Xi \perp NonDescendants_{Xi} \mid Pa_{Xi}$)

Today: The Representation Theorem

BN:  **Encodes independence assumptions**

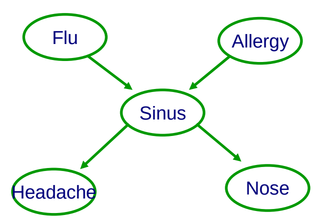
I-map: G is an I-map for P

If conditional independencies in BN are subset of conditional independencies in *P* $I_\ell(G) \subseteq I(P)$	Obtain ➡	**Joint probability distribution:** $P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$
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 ➡	**Then conditional independencies in BN are subset of conditional independencies in *P***

Factorized distributions

- Given
  - Random vars $X_1,\ldots,X_n$
  - *P* distribution over vars
  - BN structure *G* over same vars
- *P* factorizes according to *G* if

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



BN Representation Theorem – I-map to factorization

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

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

Obtain

P factorizes acc. to G

Joint probability distribution:

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

chain rule needs ordering

G* is an I-map of *P

P* factorizes according to *G

BN Representation Theorem – I-map to factorization: **Proof, part 1**

G* is an I-map of *P

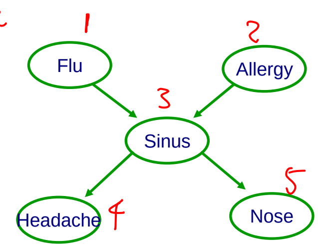
Obtain

P* factorizes according to *G

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

Topological Ordering: use this in chain rule

- Number variables such that:
  - parent has lower number than child
  - i.e., $X_i \rightarrow X_j \Rightarrow i<j$
- DAGs always have (many) topological orderings
  - find by a modification of breadth first search (not exactly what is in the book)



BN Representation Theorem – I-map to factorization: **Proof, part 2**

G* is an I-map of *P → Obtain → ***P* factorizes according to *G***

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

let's say wlog $X_1 \ldots X_n$ is a topological ordering

chain rule

$P(X_1 \ldots X_n) = P(X_1) \cdot P(X_2|X_1) \ldots P(X_i|X_1 \ldots X_{i-1}) \cdot P(X_n|X_1 \ldots X_{n-1})$

$P(X_i|X_1 \ldots X_{i-1})$

$Pa_{X_i} \subseteq \{X_1, \ldots, X_{i-1}\}$

$\{X_1, \ldots, X_{i-1}\}$ non-descendants of X_i

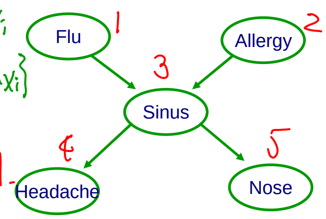
$\Rightarrow$ by local Markov $I(X_i \perp \{X_1 \ldots X_{i-1}\} / Pa_{X_i} \mid Pa_{X_i})$

$\Rightarrow P(X_i|X_1 \ldots X_{i-1}) = P(X_i|Pa_{X_i})$

q.e.d.

ALL YOU NEED:

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents
(Xi ⊥ NonDescendants$_{Xi}$ | Pa$_{Xi}$)



Adding edges doesn't hurt

- **Theorem**: Let **G** be an I-map for ***P***, any DAG **G'** that includes the same directed edges as **G** is also an I-map for ***P***.

  G' strictly more expressive than G

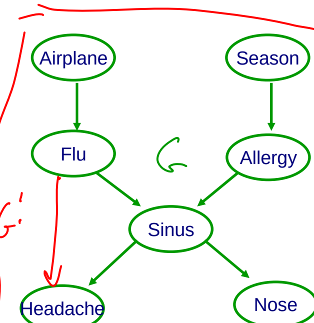
- **Proof**: G I map for P, add edge, still I-map

& P factors according to G

G' is still a DAG

$\Rightarrow$ topological ordering in G' is also one for G

use ordering $P(X_i|Pa_{X_i}(G')) = P(X_i|Pa_{X_i}(G))$



Defining a BN

- Given a set of variables $X_1 \ldots X_n$ and conditional independence assertions of P $I(P)$ tell you
- Choose an ordering on variables, e.g., $X_1, \ldots, X_n$
- For i = 1 to n who are parents of X_i
  - 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 $\mathbf{Pa}_{Xi}$ $(X_i \perp \{X_1..X_{i-1}\} \setminus \mathbf{Pa}_{X_i} \mid \mathbf{Pa}_{X_i})$
  - Define/learn CPT – $P(X_i \mid \mathbf{Pa}_{Xi})$

BN Representation Theorem – Factorization to I-map

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

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

P* factorizes according to *G

G* is an I-map of *P

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

BN Representation Theorem – Factorization to I-map: **Proof**

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 →

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

P* factorizes according to *G

G* is an I-map of *P

Homework 1!!!! ☺

The BN Representation Theorem

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

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

Obtain →

Joint probability distribution:

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

Important because:
Every *P* has at least one BN structure *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 →

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

Important because:
Read independencies of *P* from BN structure *G*

What you need to know thus far

- Independence & conditional independence
- Definition of a BN
- Local Markov assumption
- The representation theorems
  - Statement: *G* is an I-map for *P* if and only if *P* factorizes according to *G*
  - Interpretation

Announcements

- Upcoming recitation
  - Tomorrow 5 - 6:30pm in Wean 4615A
    - review BN representation, representation theorem, d-separation (coming next)
- Don't forget to register to the mailing list at:
  - https://mailman.srv.cs.cmu.edu/mailman/listinfo/10708-announce
- If you don't want to take the class for credit (will sit in or audit) – please talk with me after class

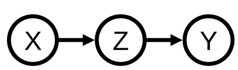
Independencies encoded in BN

- 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

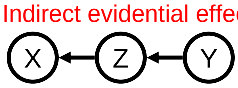
Indirect causal effect:



not nec. $(X \perp Y)$

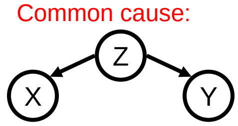
$(X \perp Y \mid Z)$

Indirect evidential effect:



$(X \perp Y \mid Z)$

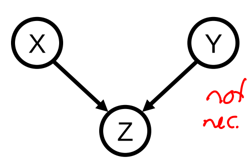
Common cause:



$(X \perp Y \mid Z)$

V-Structure

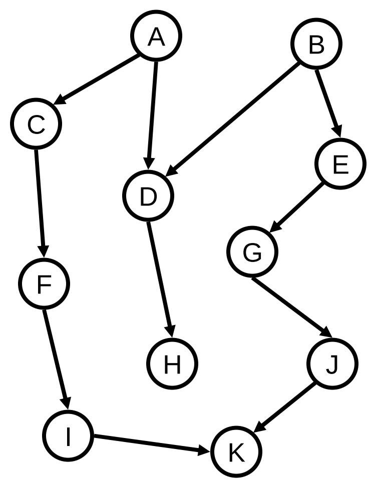
Common effect:



$(X \perp Y)$

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

Understanding independencies in BNs – Some examples



$(H \perp A \mid D)$

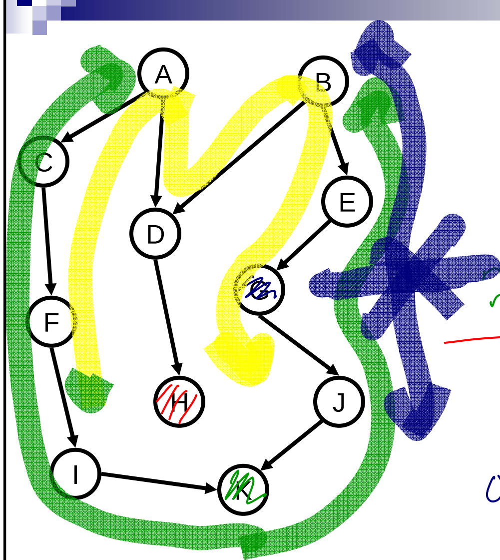
$(A \perp B)$

$(C \perp E)$

$(F \perp J)$

not nec. $(D \perp J)$

Understanding independencies in BNs – Some more examples



observe H:

not nec: $(F \perp J \mid H)$

observe K

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

Observe $\underline{G}$

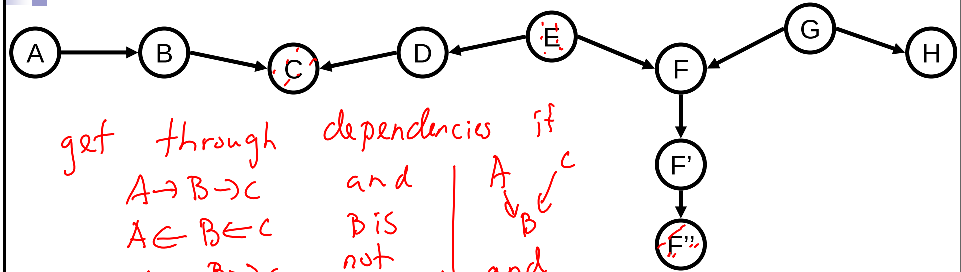
not nec. $(B \perp J)$

$(B \perp J \mid G$

Observe B

$(D \perp E \mid B)$

An active trail – Example



get through dependencies if

$A \to B \to C$

$A \leftarrow B \leftarrow C$

$A \leftarrow B \to C$

and B is not observed

$A \to B \leftarrow C$ and B or descendent of B observed

When are A and H independent?

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

$(A \perp H)$

$(A \perp H \mid C)$

not nec $(A \perp H \mid \{C, F''\})$

Active trails formalized

let "influence" flows

- A 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:
  - $X_{i-1} \to X_i \to X_{i+1}$, 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** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \to X_{i+1}$, and X_i is **not observed** ($X_i \notin \boldsymbol{O}$)
  - V-structure $X_{i-1} \to X_i \leftarrow X_{i+1}$, and X_i **is observed** ($X_i \in \boldsymbol{O}$), or **one of its descendents**

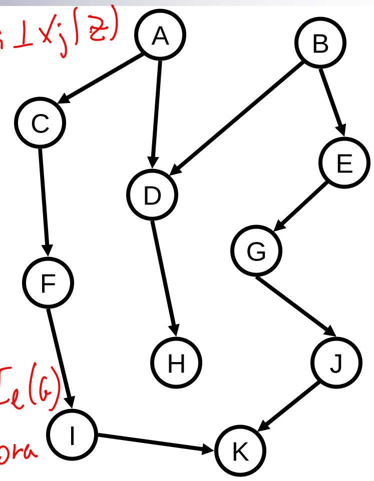
Active trails and independence?

⇒ d-sep$_G(X_i, X_j | Z)$

no active trail X_i to X_j given Z

$(X_i \perp X_j | Z)$

- **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



prove independencies beyond local Markov $I_\ell(G)$ with play with algebra

More generally: Soundness of d-separation

- Given BN structure *G*
- Set of independence assertions obtained by d-separation:
  - $I(G) = \{(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}) : \text{d-sep}_G(\mathbf{X};\mathbf{Y}|\mathbf{Z})\}$

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

local Markov d-sep

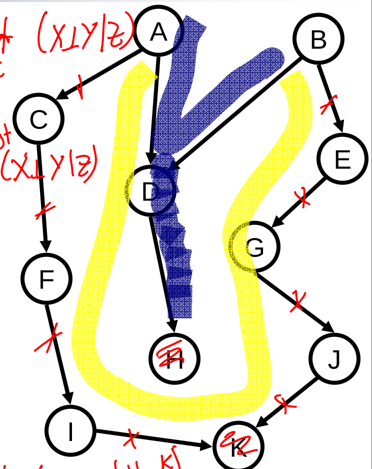
- **Theorem: Soundness of d-separation**
  - If *P* factorizes over *G* then $I(G) \subseteq I(P)$

read from graph ⇒ true for 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* ⇒ not nec $(X \perp Y \mid Z)$; $\exists P$: not $(X \perp Y \mid Z)$
- **Proof sketch**:
  - Choose an active trail between X and Y given **Z** e.g., blue trail
  - Make this trail dependent
  - Make all else uniform (independent) to avoid "canceling" out influence $(F \perp I)$ go to yellow uniform dist



not $(A \perp B \mid H, K)$ $Z = \{H, K\}$

More generally: Completeness of d-separation

- **Theorem: Completeness of d-separation**
  - For "almost all" distributions ~~that~~ 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)
- **Proof sketch:**

$X \rightarrow Y$

$P(X=x) = \theta_x$ not nec. $(X \perp Y)$

$P(Y=y \mid X=x) = \theta_{y|x}$

$P(Y=y) = \sum_x p(x) \cdot p(y|x) = \sum_x \theta_x \cdot \theta_{y|x}$

if edge unnecessary

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

$\Rightarrow \theta_{y|x} = \sum_{x'} \theta_{x'} \cdot \theta_{y|x'}$

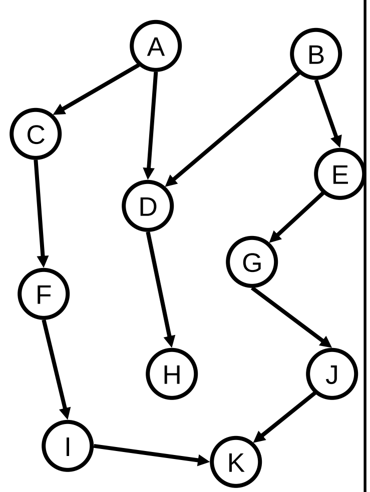
⇒ only holds for a finite set (or subspace) of parameters)

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
- 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
  - sound procedure for finding independencies $\neg$ d-sep $\Rightarrow$ dependent
  - existence of distributions with these ~~independencies~~
  - (almost) all independencies can be read directly from graph without looking at CPTs

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., $X_1, \ldots, X_n$
- 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

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!

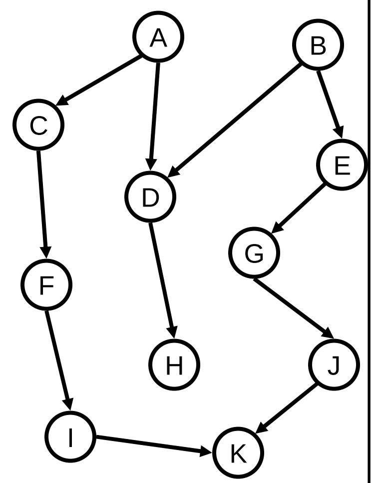
I-Equivalence

- 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?

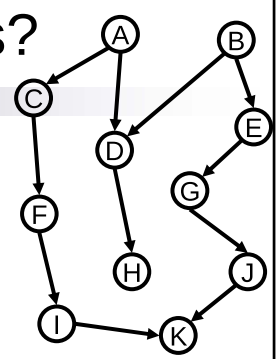
Skeleton of a BN

- **Skeleton** of a BN structure *G* is an **undirected graph** over the same variables that has an edge X–Y for every X→Y or Y→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 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 $Z \in \mathbf{U}$

- Must **not** be $X \rightarrow Z \leftarrow Y$ (not immorality):
  - When there exists **U** with $Z \in \mathbf{U}$, 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

- 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