# Drawing a BN starting from a distribution

Given a distribution $P(x_1, \ldots, x_n)$ to which we can ask any CI of the form "Is $X \perp\!\!\!\perp Y|Z$?" and get a yes/no answer. $P(X|Y,Z) = P(X|Z)$

Goal: Draw a minimal, correct BN $G$ to represent $P$.

- A DAG G is correct if all Local-CIs that are implied in G hold in P.
- A DAG G is minimal if we cannot remove any edge(s) from G and still get a correct BN for P.

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

$Y = \{0, 1, 2\}$

$X = \{0, 1\}$

$P(X|Y,Z) \neq P(X)$

$P(X=0 | Y=0) = P(X=0)$

$P(X=1 | Y=0) =$

$\vdots$

check for all

# Algorithm for drawing a BN from CIs

$x_1, \ldots, x_n =$ Choose an ordering of variables
For $i = 1 \ldots n$

- $S$=smallest subset of $Q_i = \{x_1, \ldots x_{i-1}\}$ such that $x_i \perp\!\!\!\perp Q_i - S | S$
- Make each variable in $S$ a parent of $x_i$

# Examples

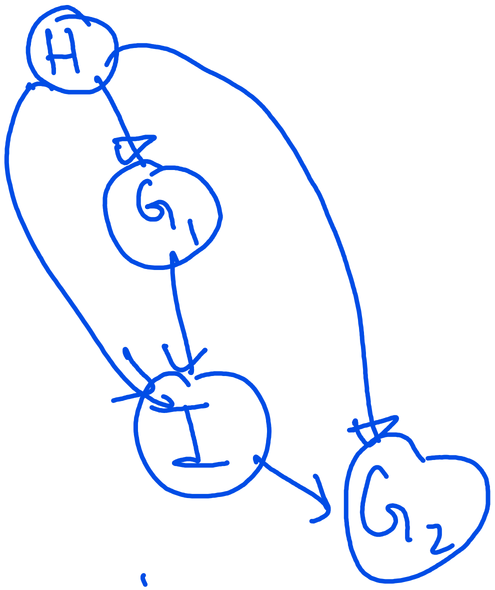
$H \equiv$ hardworking or not

$I \equiv$ intelligent or not

$G_1 \equiv$ Grade in course 1

$G_2 =$ " " " 2

$H, G_1, I, G_2$




(1) $G_1 \perp\!\!\!\perp H / S = \phi \rightarrow$ ✗

(2) $G_1 \perp\!\!\!\perp \phi \mid \underline{S = H}$ ✓

for I

$I \perp\!\!\!\perp \{G_1, H\} \mid \phi$ ✗

$I \perp\!\!\!\perp G_1 \mid S = H$ ✗

$I \perp\!\!\!\perp H \mid S = G_1$ ✗

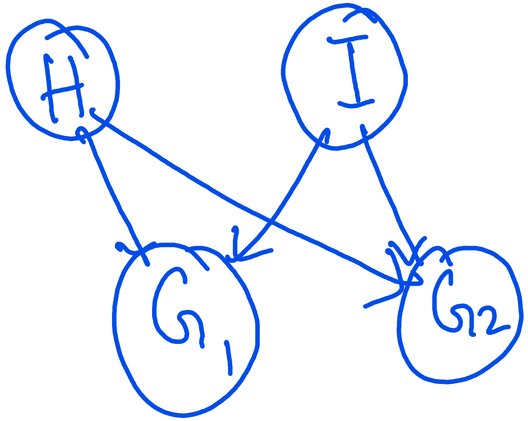
$I \perp\!\!\!\perp \{\phi\} \mid H, G_1$

---

$G_2 \perp\!\!\!\perp G_1 \mid H, I$ ✓

# Examples

Alternative ordering $H, I, G_1, G_2$



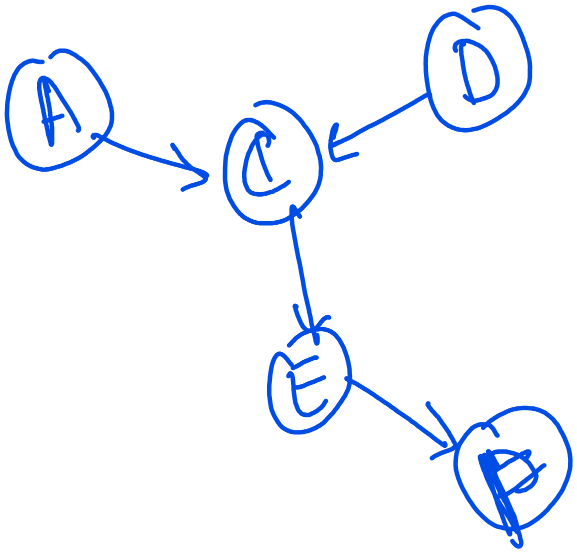
# Order is important

# CIs and Fractorization I

## Theorem

*Given a distribution $P(x_1, \ldots, x_n)$ and a DAG $G$, if $P$ satisfies Local-CI induced by $G$, then $P$ can be factorized as per the graph.*
*Local-CI$(P, G) \implies$ Factorize$(P, G)$*

Definition: Topological order of nodes in a DAG:




D, A, C, E, P
A, D, C, E, P
Topological ordered

D, P, C, D, E  Not topol.

# CIs and Fractorization II

## Proof.

- $x_1, x_2, \ldots, x_n$ topographically ordered (parents before children) in $G$.
- For every $x_i$ the set $x_1, \ldots, x_{i-1}$ are all non-descents i.e. $x_1, \ldots, x_{i-1} = Pa_G(x_i) \cup ND'(x_i)$
- Local CI($P, G$): $P(x_i|x_1, \ldots, x_{i-1}) = P(x_i|Pa_G(x_i) \cup ND'(x_i)) = P(x_i|Pa_G(x_i))$
- Chain rule: $P(x_1, \ldots, x_n) = \prod_i P(x_i|x_1, \ldots, x_{i-1}) = \prod_i P(x_i|Pa_G(x_i))$
- $\implies$ Factorize($P, G$)

□

Also as Theorem 3.1 in KF book.

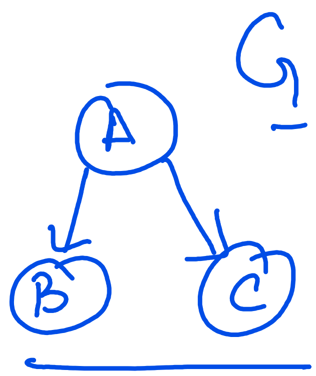
# CIs and Fractorization

## Theorem

*Given a distribution* $P(x_1, \ldots, x_n)$ *and a DAG* $G$, *if* $P$ *can be factorized as per* $G$ *then* $P$ *satisfies Local-CI induced by* $G$.
*Factorize*$(P, G) \implies$ *Local-CI*$(P, G)$

Proof skipped. (Refer Theorem 3.2 in KF book.)

$P(A, B, C) = P(B|A)\,P(C|A)\,P(A)$




Factorize$(P, G)$ holds

$\Rightarrow$ Local-CIs of $G$: $B \perp\!\!\!\perp C \mid A$

$C \perp\!\!\!\perp B$

# What have we achieved so far?

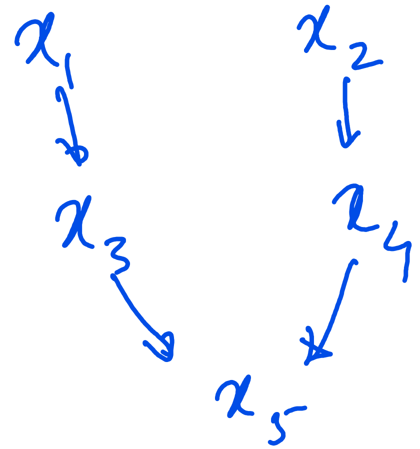
- When a large number of variables are pairwise correlated, we may still be able to represent their joint distribution efficiently by exploring conditional independencies
- BNs provide a principled procedure of achieving this goal: Using CIs draw a BN, then express the joint distribution as a product of factors involving a node and only its immediate parents.
- Many real-life problems can be expressed as BNs.

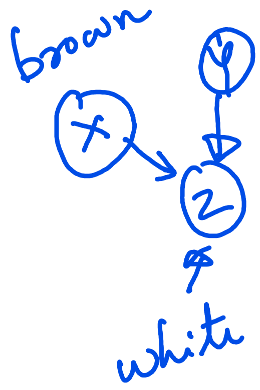
# Examples of CIs that hold in BN but not covered by local-CI

CI properties: $X \perp\!\!\!\perp Y \not\Rightarrow X \perp\!\!\!\perp Y \mid Z.$

but if $X \perp\!\!\!\perp Z$ then $X \perp\!\!\!\perp Y \mid Z$

1. $X \perp\!\!\!\perp Y, W \mid Z \rightarrow X \perp\!\!\!\perp Y \mid Z, W$ $\Longleftrightarrow$
2. $\underline{X \perp\!\!\!\perp W \mid Y, X \perp\!\!\!\perp Y} \rightarrow X \perp\!\!\!\perp Y, W$




$$\left.\begin{array}{l} x_3 \perp\!\!\!\perp x_4, x_2 \mid x_1 \\ x_4 \perp\!\!\!\perp x_3, x_1 \mid x_2 \\ x_1 \perp\!\!\!\perp x_2 \end{array}\right\} \longrightarrow \underline{x_3 \perp\!\!\!\perp x_4}$$

# Global CIs in a BN I

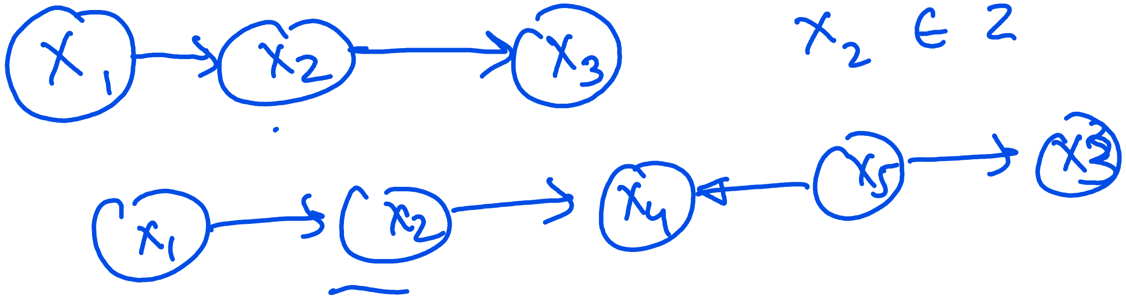
Three sets of variables $X, Y, Z$. If $Z$ d-separates $X$ from $Y$ in BN then, $X \perp\!\!\!\perp Y | Z$.

In a directed graph $H$, $Z$ d-separates $X$ from $Y$ if all paths $P$ from any $X$ to $Y$ is blocked by $Z$.

A path $P$ is blocked by $Z$ when any of these 4 conditions hold:

1. $\overset{X}{x_1} \leftrightharpoons x_2 \leftrightharpoons \dots x_{i-1} \rightarrow x_i \rightarrow x_{i+1} \dots \leftrightharpoons \overset{Y}{x_k}$ and $x_i \in Z$

$x_2 \in Z$ then $x_1 \perp\!\!\!\perp x_3 | Z$




$X_1 \perp\!\!\!\perp X_3 | X_2$

# Global CIs in a BN II

2. $x_1 \leftrightharpoons x_2 \leftrightharpoons \ldots x_{i-1} \leftarrow x_i \leftarrow x_{i+1} \ldots \leftrightharpoons x_k$ and $x_i \in Z$

3. $x_1 \leftrightharpoons x_2 \leftrightharpoons \ldots x_{i-1} \leftarrow x_i \rightarrow x_{i+1} \ldots \leftrightharpoons x_k$ and $x_i \in Z$

   $x_1 \leftarrow x_2 \rightarrow x_3$ $\qquad$ $x_1 \perp\!\!\!\perp x_3 \mid x_2$

4. $x_1 \leftrightharpoons x_2 \leftrightharpoons \ldots x_{i-1} \rightarrow x_i \leftarrow x_{i+1} \ldots \leftrightharpoons x_k$ and $x_i \notin Z$ and $Desc(x_i) \notin Z$

# Global CIs in a BN III

$$x_1 \rightarrow x_2 \leftarrow x_3$$

$Z = \phi$

(1) $X_1 \perp\!\!\!\perp X_3$ ✓

$X_1 \not\perp\!\!\!\perp X_3 \mid X_2$

$Z = \{x_2\}$

$X = \{x_2\}$

$Y = \{x_1\}$

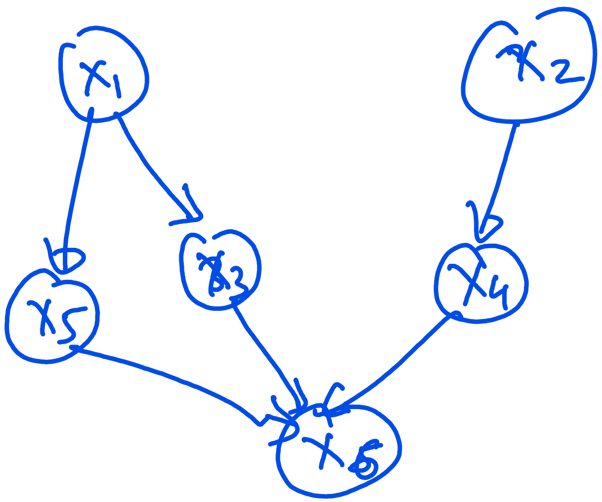
$P = x_2 \leftarrow x_1$

# Global CIs in a BN

## Theorem

*The d-separation test identifies the* complete *set of conditional independencies that hold in* all *distributions that conform to a given Bayesian network.*

# Global CIs Examples




$X \equiv \{x_3, x_5\}$

$Z = \phi$

$Y = \{x_4\}$

$X \perp\!\!\!\perp Y$

(1) $x_3 \rightarrow x_6 \leftarrow x_4$ blocked

(2) $x_5 \rightarrow x_6 \leftarrow x_4$ blocked

3

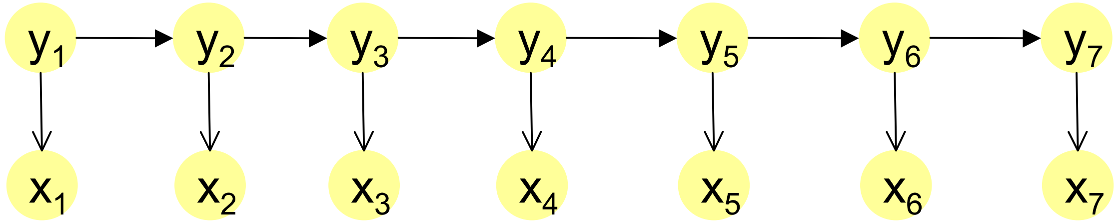
$\{x_3, x_5\} \perp\!\!\!\perp x_4$

# Global CIs and Local-CIs

In a BN, the set of CIs combined with the axioms of probability can be used to derive the Global-CIs.
Proof is long but easy to understand. Sketch of a proof available in the supplementary.

# Popular Bayesian networks

- Hidden Markov Models: speech recognition, information extraction




  - State variables: discrete phoneme, entity tag
  - Observation variables: continuous (speech waveform), discrete (Word)
- Kalman Filters: State variables: continuous
  - Discussed later
- Topic models for text data
  1. Principled mechanism to categorize multi-labeled text documents while incorporating priors in a flexible generative framework
  2. Application: news tracking
- QMR (Quick Medical Reference) system
- PRMs: Probabilistic relational networks: