Graphical Model Basics

Cyrill Stachniss

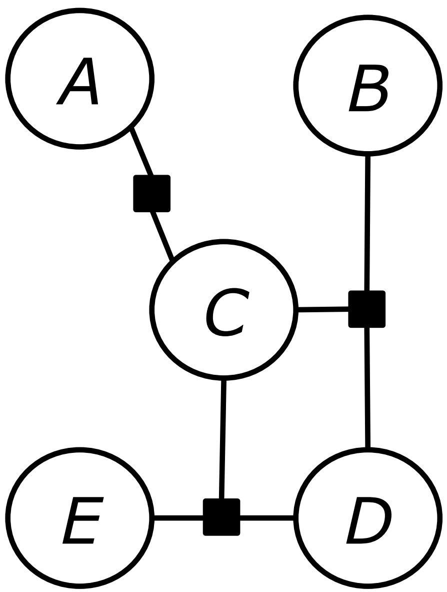
This lecture is strongly influenced by Zoubin Ghahramani's GM tutorials

Probabilistic Graphical Models

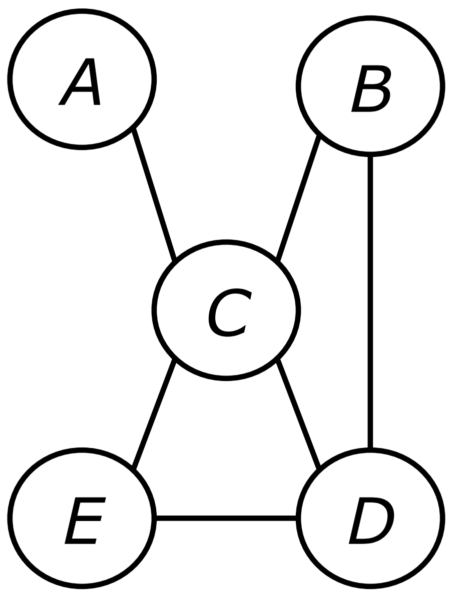
- Marriage between probability theory and graph theory
- Tool for dealing with uncertainty, independence, and complexity
- Notion of modularity – a complex system that consists of simpler parts
- Probability theory is the “glue” for the individual parts
- Play an increasingly important role in robotics, vision, and machine learning

Three Typical Graphical Models

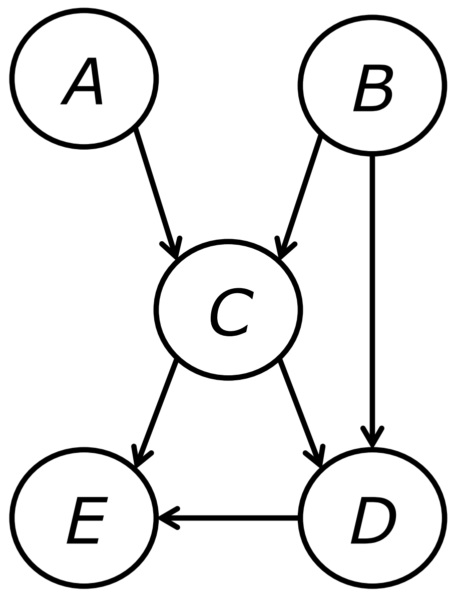
- Nodes represent random variables
- Edges represent statistical dependencies between these variables



factor graph



undirected graph



directed graph

Why Graphical Models?

- Abstract out the conditional independence relationships between the variables from their parametric forms
- Directly answer questions: “Is A dependent on B given we know the value of C?”
- Allow to realize efficient message passing algorithms to do inference:
  “What is $P(A \mid C = c)$ without enumerating all combination of variables?”

Conditional Independence

- Conditional Independence

$$X \perp\!\!\!\perp Y \mid V \quad \Leftrightarrow \quad p(X \mid Y, V) = P(X \mid V)$$

- as well as

$$X \perp\!\!\!\perp Y \mid V \quad \Leftrightarrow \quad p(X, Y \mid V) = P(X \mid V)P(Y \mid V)$$

- Conditional independence is different to the "normal"/marginal independence

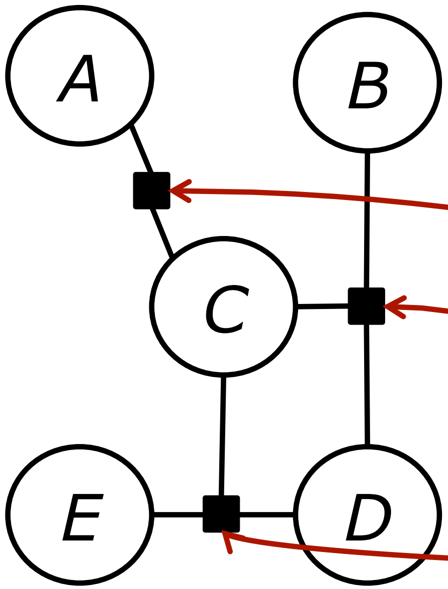
$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad X \perp\!\!\!\perp Y \mid \emptyset \quad \Leftrightarrow \quad p(X, Y) = P(X)P(Y)$$

Conditional Independence Examples

- Speeding fine $\perp\!\!\!\perp$ type of car $|$ speed
- Lung cancer $\perp\!\!\!\perp$ yellow teeth $|$ smoker
- Ability of team A $\perp\!\!\!\perp$ ability of Team B
- NOT (ability of team A $\perp\!\!\!\perp$ ability of Team B $|$ outcome of A vs. B game)

Factor Graphs

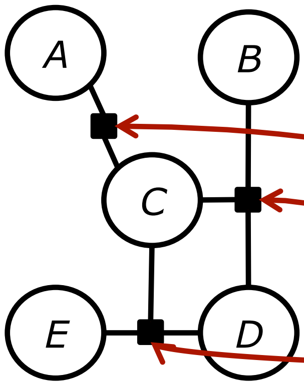
- Two types of nodes
  - Circles are random variables
  - Boxes are factors in the joint distribution



$$P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C,D)\, g_3(C,D,E)$$

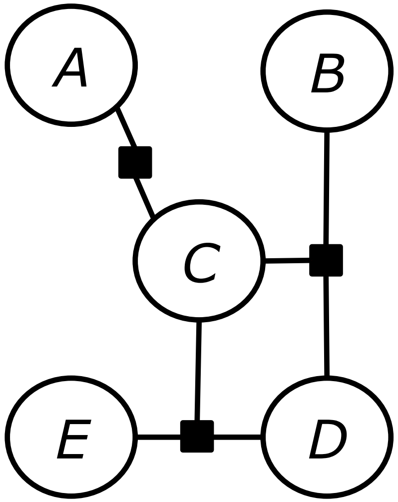
Factors

- Factors (g_i) are non-negative functions
- Two nodes are **neighbors** if they share a common factor (g_i)
- The normalizer $\frac{1}{Z}$ ensures that the sum/ integral over the joint equals to 1
- E.g. $Z = \int_a \int_b \int_c \int_d \int_e g_1(a,c) g_2(b,c,d) g_3(c,d,e)$

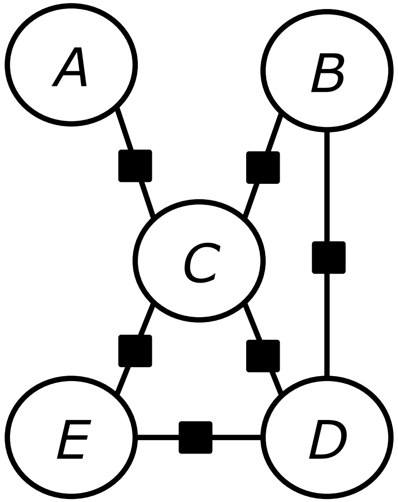


$$P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C,D)\, g_3(C,D,E)$$

Two Different Factor Graphs



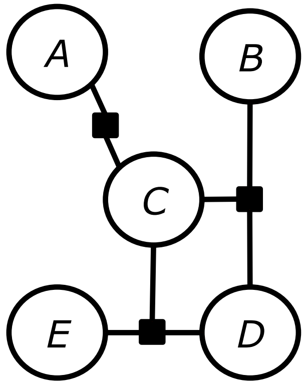
$$P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C,D)\, g_3(C,D,E)$$



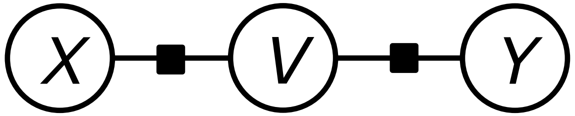
$$P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C)\, g_3(B,D)\, g_4(C,D)\, g_5(C,E)\, g_6(D,E)$$

Factor Graphs: Definitions and Facts

- Two nodes are **neighbors** if they share a common factor
- A **path** is a sequence of neighboring nodes
- $X \perp\!\!\!\perp Y \mid \mathcal{V}$ if every path between X and Y contains some node $V \in \mathcal{V}$
- E.g., $B \perp\!\!\!\perp E \mid C, D$
- Given the neighbors of X $(\mathrm{ne}(X))$, the variable X is **conditionally independent** of all other variables



Proving Cond. Independence (Idea)

- Show: $X \perp\!\!\!\perp Y \mid V$ given the FC 

- CI: $X \perp\!\!\!\perp Y \mid V \quad \Leftrightarrow \quad p(X \mid Y, V) = P(X \mid V)$
- Factor graph: $P(X, Y, V) = \frac{1}{Z} g_1(X, V) g_2(Y, V)$
- Summing over X:

$$P(Y, V) = \sum_x P(x, Y, V) = \frac{1}{Z} \left(\sum_x g_1(x, V) \right) g_2(Y, V)$$

- Dividing results in

$$P(X \mid Y, V) = \boxed{\frac{g_1(X, V)}{\sum_x g_1(x, V)}}$$

not dependent on Y

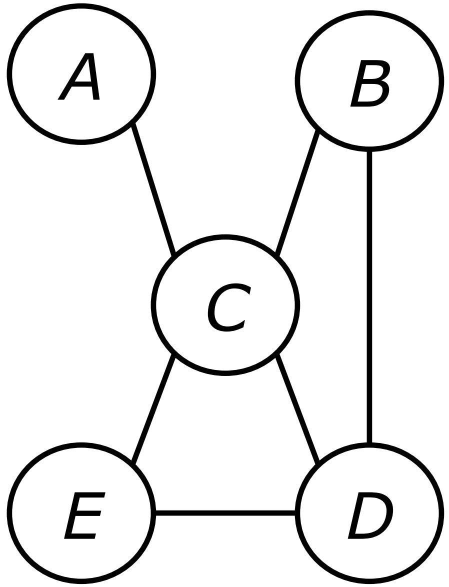
- The r.h.s. does not depend on Y (but on V)
- The factorization implies cond. independence

■

Undirected Graphs

- Also called **Markov networks** or **MRFs**
- Defines a joint probability distribution based on the cliques of the graph

$$P(X) = \frac{1}{Z} \prod_j g_j(X_{C_j})$$

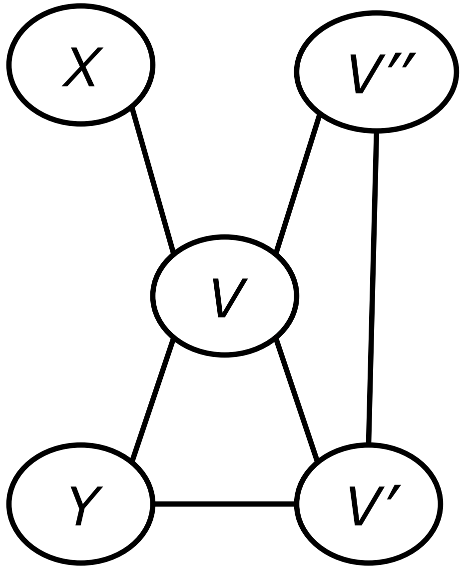


- where C_j are the cliques (fully connected subgraphs)
- Very similar to factor graphs!
- Here:

$$P(A, \ldots, E) = \frac{1}{Z} g_1(A, C) g_2(B, C, D) g_3(C, D, E)$$

Cond. Independence

- $X \perp\!\!\!\perp Y \mid \mathcal{V}$ if every path between X and Y contains some node $V \in \mathcal{V}$
- $\mathcal{V}$ is a **Markov Blanket** for X iff $X \perp\!\!\!\perp Y \mid \mathcal{V}$ for all $Y \notin \{X \cup \mathcal{V}\}$
- A minimal Markov Blanket is a Markov Boundary

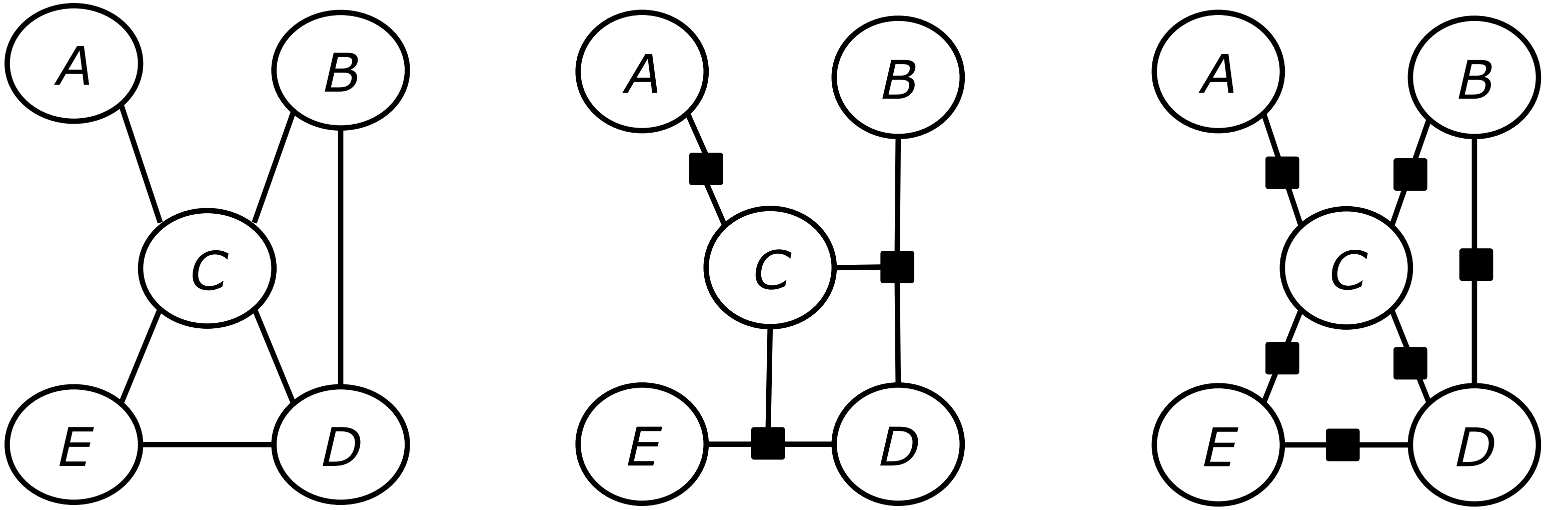


- For X, we have here
  - Markov Blanket(s): $\{V\}, \{V, V'\}, \{V, V''\}, \{V, Y\}, \ldots$
  - Markov Boundary: $\{V\}$

Markov Boundary

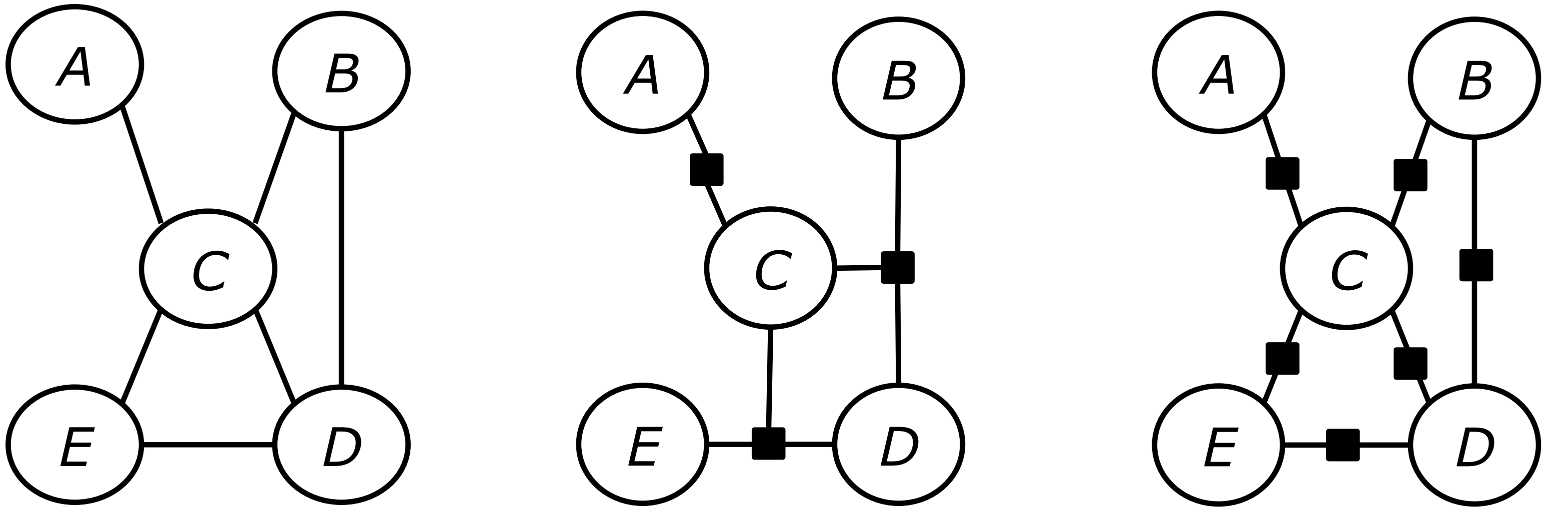
- A minimal Markov Blanket is a Markov Boundary
- For factor graphs and undirected graphical models, the Markov Boundary of X = $\mathrm{ne}(X)$

Factors Graphs vs. Undirected G.



- The nodes always have the same neighbors
- All graphs encode the same conditional independence relations

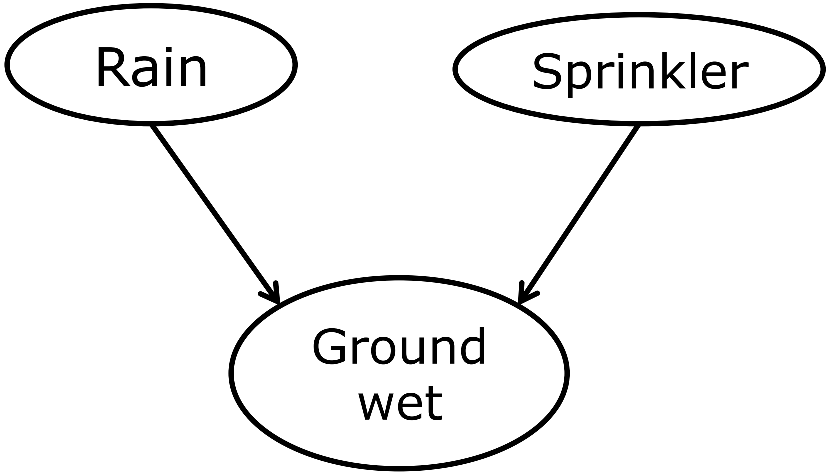
Factors Graphs vs. Undirected G.

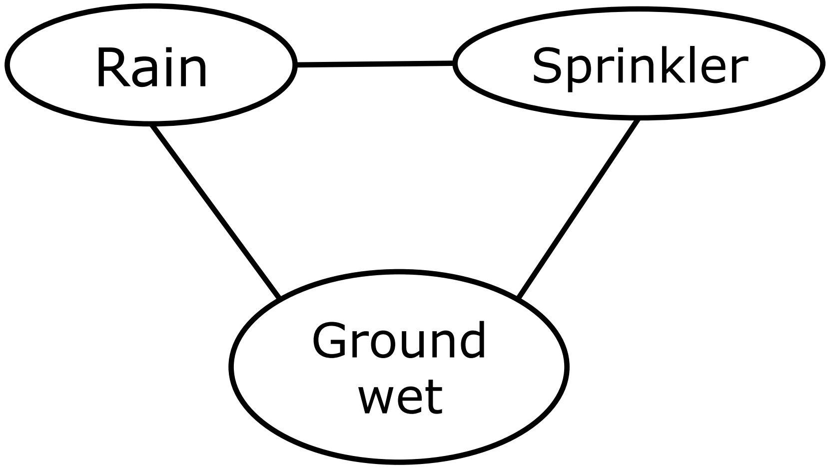


- The nodes always have the same neighbors
- All graphs encode the same conditional independence relations
- #3: encodes that only pairwise factors are needed to model the joint distribution
- Different complexities, cubic vs. quadratic!

Problems and Limitations

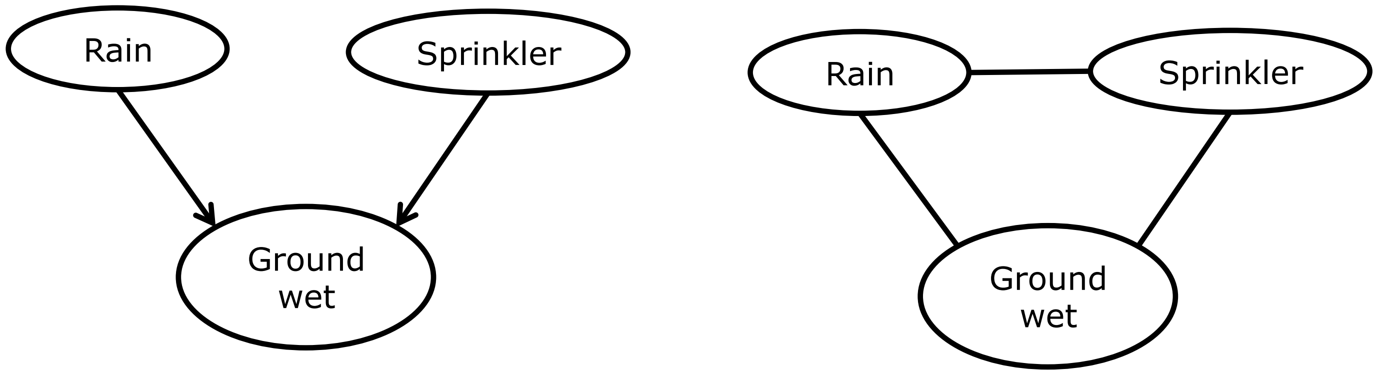
- Often two variables are connected because some other variable depends on them
- Many independencies are underrepresented





- Difference between **independence** and **conditional independence**

Explaining Away

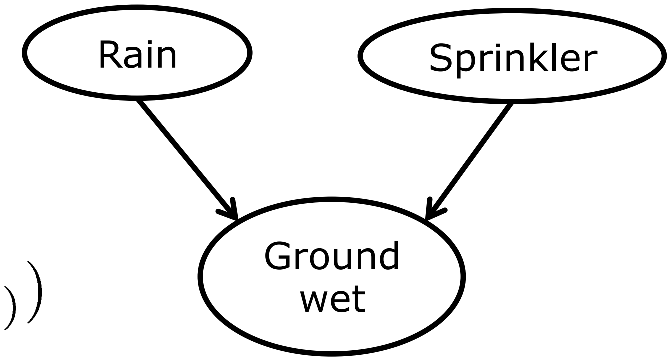


- Rain and sprinkler are independent (given nothing), but conditionally dependent given the ground is wet
- **"Explaining Away"**: Observing that the sprinkler is on, would explain away the observation that the ground was wet, making it less probable that it rained.

Directed Acyclic Graphs

- Also called Bayesian networks
- Arrows represent "influences"
- Results in the joint distribution

$$P(R, S, G) = P(R)P(S)P(G \mid R, S)$$

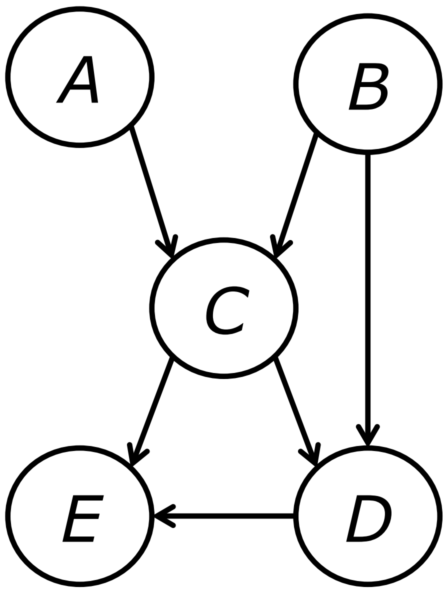


- In general

$$P(X_1, \ldots, X_n) = \prod_i p(X_i \mid X_{\text{parents}(i)})$$

Example

- In general $P(X_1, \ldots, X_n) = \prod_i p(X_i \mid X_{\text{parents}(i)})$



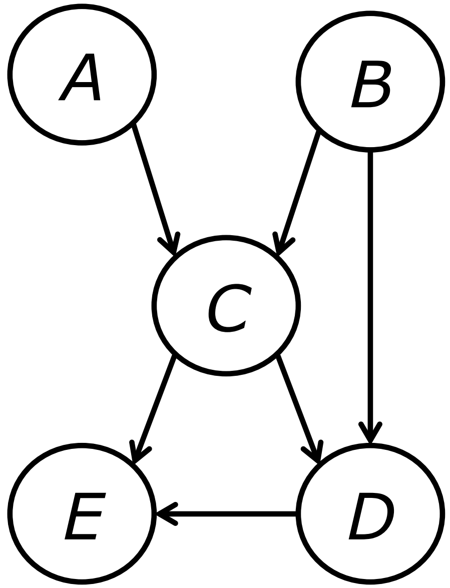
$$
\begin{gathered}
P(A, B, C, D, E) = \\
P(A)P(B) \\
P(C \mid A, B) \\
P(D \mid B, C) \\
P(E \mid C, D)
\end{gathered}
$$

D-Separation and Cond. Independence

- $X \perp\!\!\!\perp Y \mid \mathcal{V}$ if $\mathcal{V}$ **d-separates** X from Y
- $\mathcal{V}$ d-separates X from Y if every undirected path between X and Y is blocked by $\mathcal{V}$
- A path is **blocked** by $\mathcal{V}$ if there is a node W on the graph such that either:
  - W has converging arrows along the path ($\rightarrow$ W $\leftarrow$) and neither W nor its descendants are observed (in $\mathcal{V}$), or
  - W does not have converging arrows along the path ($\rightarrow$ W $\rightarrow$ or $\leftarrow$ W $\rightarrow$) and W is observed (W $\in \mathcal{V}$).

D-Separation Examples

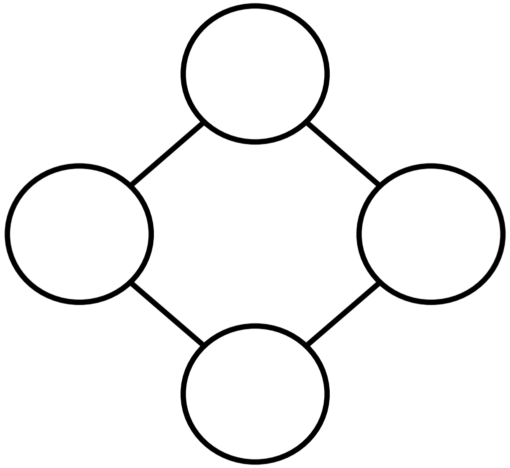
- $A \perp\!\!\!\perp B$
- not $A \perp\!\!\!\perp B \mid C$
- $A \perp\!\!\!\perp D \mid B, C$
- not $A \perp\!\!\!\perp B \mid E$



- For Bayesian networks: Markov Boundary of X = $\mathrm{parents}(X) \cup \mathrm{children}(X) \cup \mathrm{parents\ of\ children}(X)$

Expressiveness (1)

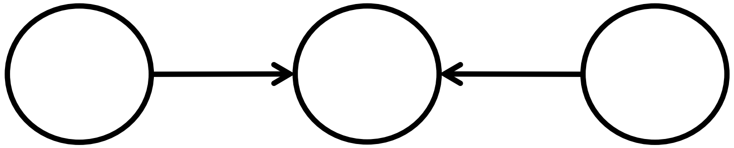
- No directed acyclic graph can represent these any only these independencies



- No matter how we direct the arrows there will always be two non-adjacent parents sharing a common child. This implies dependence in the directed graph but independence in undirected graph

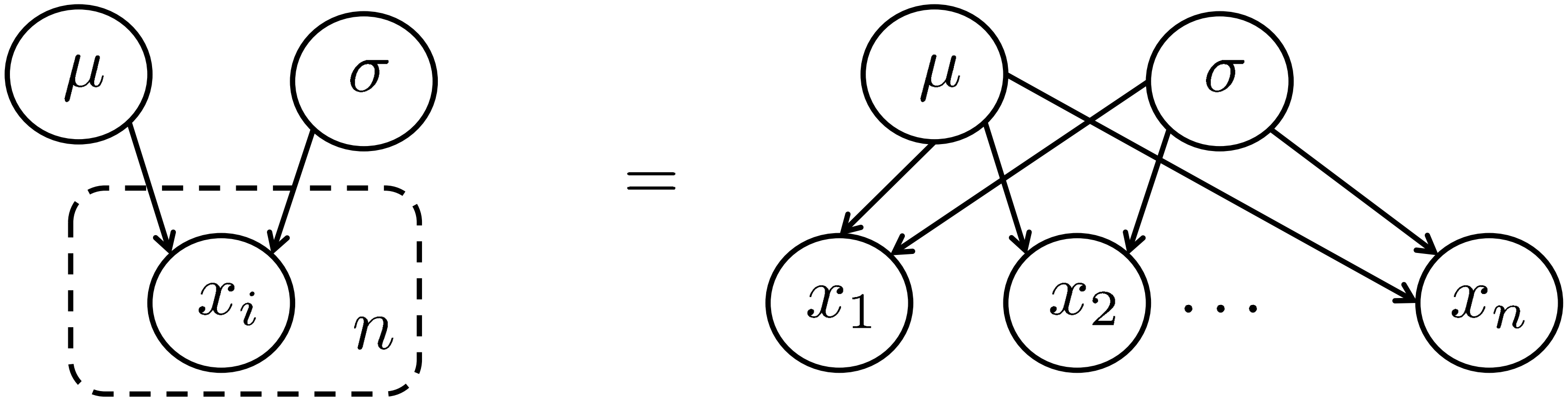
Expressiveness (2)

- No undirected graph of factor graph can represent these any only these independencies



- Directed graphs are better at expressing causal generative models, undirected graphs are better at representing soft constraints between variables.

Plate Notation



- A set of n random points generated by a Gaussian

$$p(x_1, \ldots, x_n, \mu, \sigma) = p(\mu) p(\sigma) \prod_{i=1}^{n} p(x_i \mid \mu, \sigma)$$

Summary

- Three typical graphical models
  - Factor graphs
  - Undirected graphs / Markov networks
  - Directed graphs / Bayesian networks
- Independence and conditional independence
- Markov boundaries and d-separation
- Directed vs. undirected graphs
  - Directed graphs are better at expressing causal generative models
  - Undirected graphs are better at representing soft constraints between variables.