Machine Learning Department
School of Computer Science
Carnegie Mellon University

Undirected Graphical Models

Matt Gormley
Lecture 3
Feb. 8, 2021

Q&A

Q: How will I earn the 5% participation points?

A: Very gradually. There will be a few aspects of the course (in-class polls, out-of-class polls, surveys, meetings with the course staff) that we will attach participation points to.

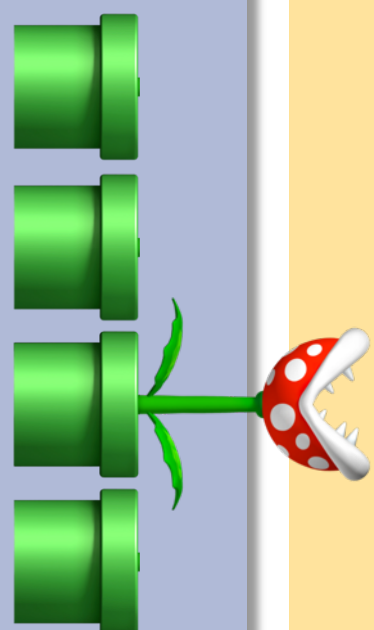
That said, we might not actually use the whole 5% that is being held out.

First In-Class Poll

Question:

How are you participating in class today?

A. Laptop
B. Smart phone
C. Pay phone
D. Desktop

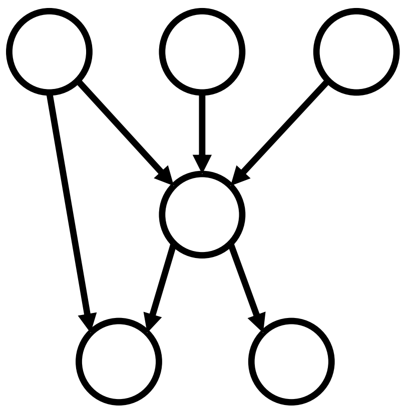


Answer:

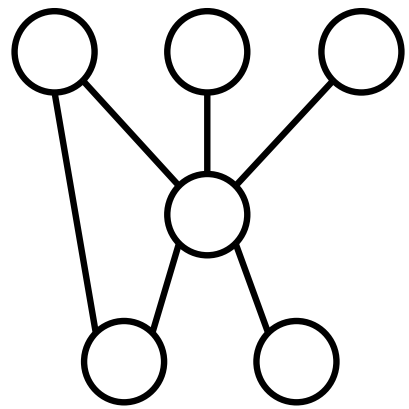
TYPES OF GRAPHICAL MODELS

Three Types of Graphical Models

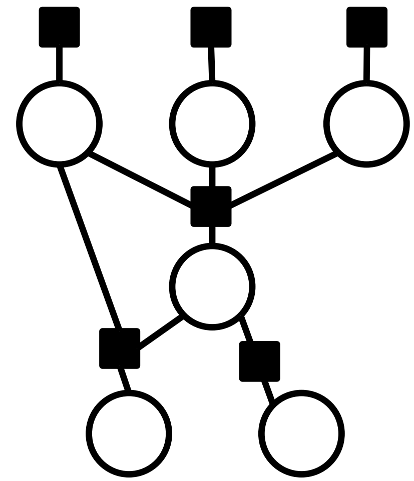
Directed Graphical Model



Undirected Graphical Model



Factor Graph



Key Concepts for Graphical Models

Graphical Models in General

1. A graphical model defines a **family of probability distributions**
2. That family shares in common a set of **conditional independence assumptions**
3. By choosing a **parameterization** of the graphical model, we obtain a single **model** from the family
4. The model may be either **locally or globally normalized**

Ex: Directed G.M.

1. Family:
2. Conditional Independencies:
3. Example parameterization:
4. Normalization:

Key Concepts for Graphical Models

Graphical Models in General

1. A graphical model defines a **family of probability distributions**
2. That family shares in common a set of **conditional independence assumptions**
3. By choosing a **parameterization** of the graphical model, we obtain a single **model** from the family
4. The model may be either **locally or globally normalized**

Ex: Undirected G.M.

1. Family:
2. Conditional Independencies:
3. Example parameterization:
4. Normalization:

Key Concepts for Graphical Models

Graphical Models in General

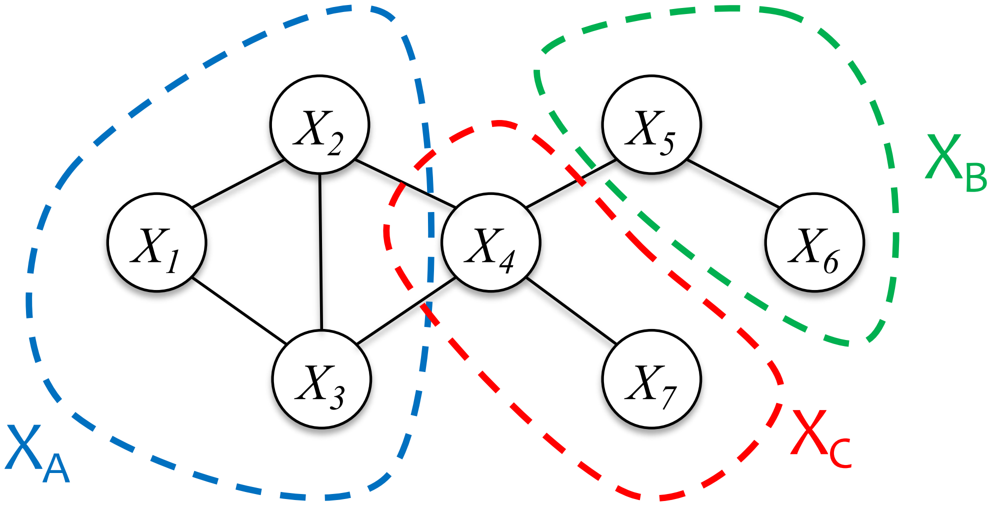
1. A graphical model defines a **family of probability distributions**
2. That family shares in common a set of **conditional independence assumptions**
3. By choosing a **parameterization** of the graphical model, we obtain a single **model** from the family
4. The model may be either **locally or globally normalized**

Ex: Factor Graph

1. Family:
2. Conditional Independencies:
3. Example parameterization:
4. Normalization:

UNDIRECTED GRAPHICAL MODELS

Undirected Graphical Models

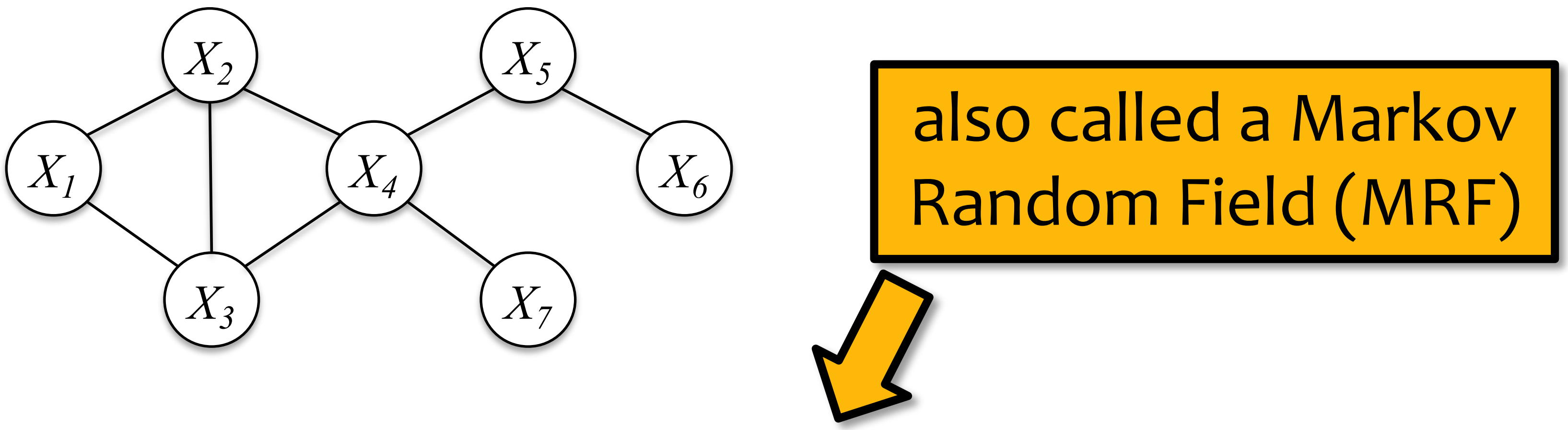


Notation: Let X_S denote all the variables with indices in the set $S \subset \mathbb{Z}^+$

Undirected Graph Terminology

- *Definition*: a **clique** is a set of fully connected nodes (e.g. $\{X_1, X_2\}$ or $\{X_1, X_2, X_3\}$)
- *Definition*: a **maximal clique** is a clique to which adding any node makes it no longer a clique (e.g. $\{X_1, X_2, X_3\}$ but not $\{X_1, X_2\}$)
- *Definition*: a set of nodes X_C **separates** sets X_A and X_B if removing X_C leaves no path from a node in X_A to one in X_B. (e.g. $\{X_4, X_7\}$ separates $\{X_1, X_2, X_3\}$ and $\{X_5, X_6\}$)

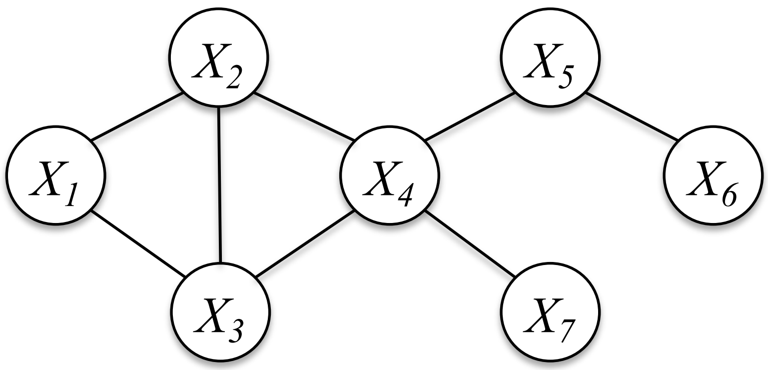
Undirected Graphical Models



Def: an **undirected graphical model (UGM)** consists of a graph G (qualitative specification) and potential functions ψ (quantitative specification)

- The graph G is an undirected graph over random variables $X_1, \ldots, X_T$ and cycles are permitted
- The potential functions ψ, also called "factors", are used to define the joint probability

Undirected Graphical Models



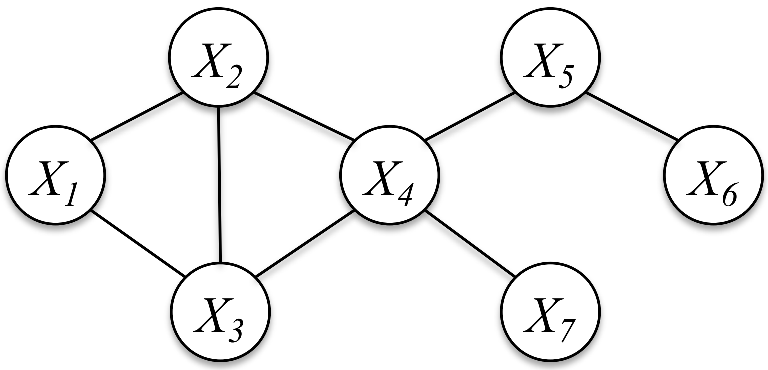
1. we have one potential function (aka. factor) per clique
2. potential functions must be non-negative
   $$\psi_C(x_C) \geq 0, \; \forall C, x_C$$
3. Z is the partition function → globally normalized model
   $$Z = \sum_{\mathbf{x} \in \mathcal{X}} \prod_{C \in \mathcal{C}} \psi_C(X_c)$$
   $$= \sum_{\mathbf{x} \in \mathcal{X}} s(\mathbf{x})$$

***Def*: Joint probability of a UGM**

$$p(x_1, \ldots, x_T) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

where $\mathcal{C}$ is the set of all cliques and $C \in \mathcal{C}$ is an index set $\Rightarrow C \subseteq \{1, \ldots, T\}$

Undirected Graphical Models



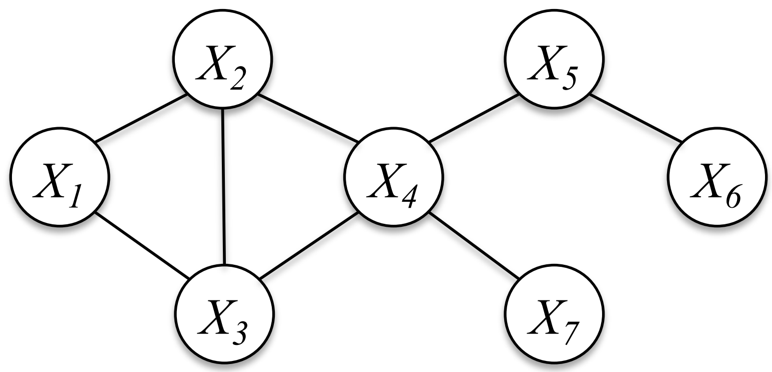
<u>Def</u>: A distribution is said to **factor according to the graph** G if it can be written as

$$p(x_1, \ldots, x_T) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

where $\mathcal{C}$ is the set of all cliques and
$C \in \mathcal{C}$ is an index set $\Rightarrow C \subseteq \{1, \ldots, T\}$

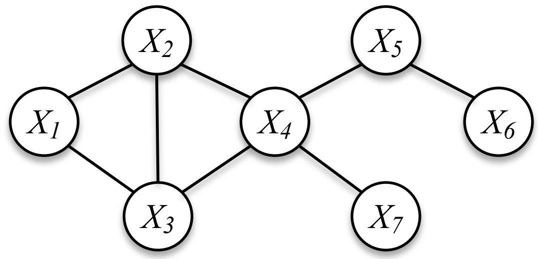
Undirected Graphical Models

Ex: Joint probability of UGM



$$p(\mathbf{x}) = \frac{1}{Z}\Big(\psi_{1,2,3}(x_1, x_2, x_3)\, \psi_{2,3,4}(x_2, x_3, x_4)$$
$$\psi_{4,5}(x_4, x_5)\, \psi_{5,6}(x_5, x_6)\, \psi_{4,7}(x_4, x_7)\Big)$$

Potential Functions for UGM



$$p(\mathbf{x}) = \frac{1}{Z}\Big(\psi_{1,2,3}(x_1, x_2, x_3)\,\psi_{2,3,4}(x_2, x_3, x_4)\,\psi_{4,5}(x_4, x_5)\,\psi_{5,6}(x_5, x_6)\,\psi_{4,7}(x_4, x_7)\Big)$$

How should we **interpret** the potential functions in a UGM?

- *Idea #1*: Maybe as **marginals** of the distribution? In general, no.

$$p(\mathbf{x}) \neq \frac{1}{Z}(p(x_1, x_2, x_3)p(x_2, x_3, x_4)p(x_4, x_5)p(x_5, x_6)p(x_4, x_7))$$

- *Idea #2*: Maybe as **conditionals** of the distribution? In general, no.

$$p(\mathbf{x}) \neq \frac{1}{Z}(p(x_1|x_2, x_3)p(x_2, x_3|x_4)p(x_4|x_5)p(x_5|x_6)p(x_7|x_4))$$

Potential Functions for UGM

Whiteboard

- Simple example of potential functions as tables

Compactness of a UGM

Consider random variables $X_1, X_2, \ldots, X_T$ where $X_i \in \mathcal{X}$, where $|\mathcal{X}| = R$

- To represent an arbitrary distribution $P(\mathbf{X})$ via a single joint probability table requires $R^T - 1$ values

Exponential in T

- If the distribution factors according to a graph G and $\max_{C \in \mathcal{C}} |C| \leq D$

  then each $\psi_C(X_C)$ needs only R^D values for a total of only $T(R^D)$ values

Linear in T

Compactness of BayesNet

Question:

Suppose we have a DGM over T variables ranging over R values each. The distribution factors according to a graph G where each node has at most D parents.

How many parameters are needed to represent the distribution?

A. $T^D (R^{D+1} - 1)$

B. $T(R^{D+1} - 1)$

C. $T^D (R^D(R - 1))$

D. $T(R^D(R - 1))$

E. TDR

Answer:

CONDITIONAL INDEPENDENCIES OF UGMS

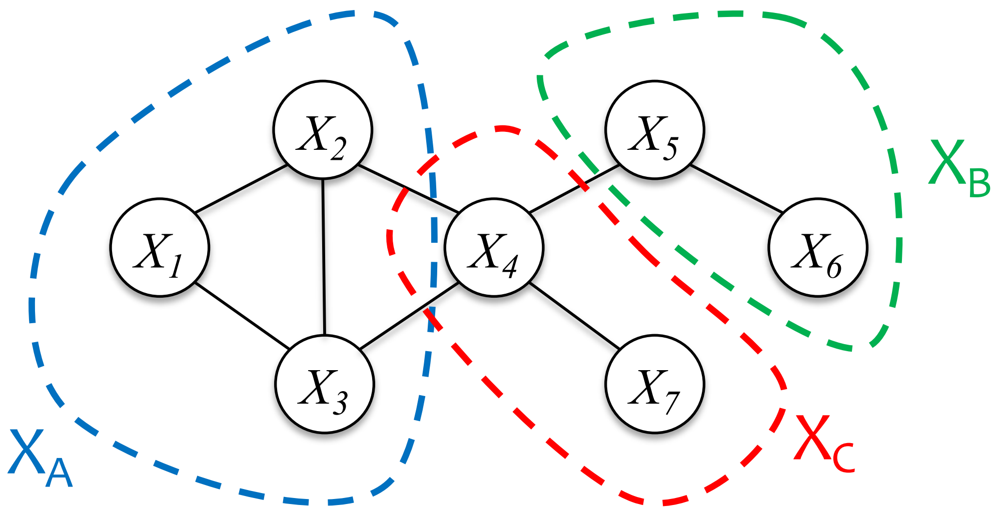
Undirected Graphical Models

Conditional Independence Semantics

Consider a distribution over r.v.s $X_1, \ldots, X_T$

For a UGM and any disjoint index sets A, B, C,
(i.e., $A \subseteq \{1, \ldots, T\}$, $B \subseteq \{1, \ldots, T\}$, $C \subseteq \{1, \ldots, T\}$)

X_A is **conditionally independent** of X_B given X_C iff X_C separates sets X_A and X_B



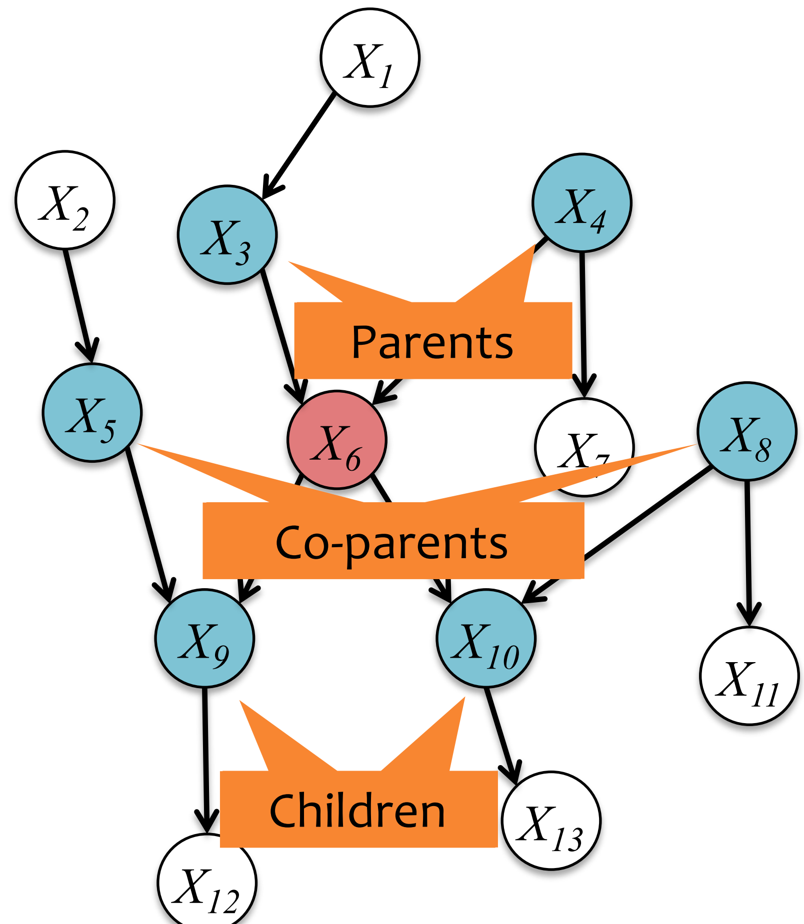
Markov Blanket (Directed)

Def: the **co-parents** of a node are the parents of its children

Def: the **Markov Blanket** of a node in a **directed** graphical model is the set containing the node's parents, children, and co-parents.

Theorem: a node is **conditionally independent** of every other node in the graph given its **Markov blanket**

Example: The Markov Blanket of X_6 is $\{X_3, X_4, X_5, X_8, X_9, X_{10}\}$

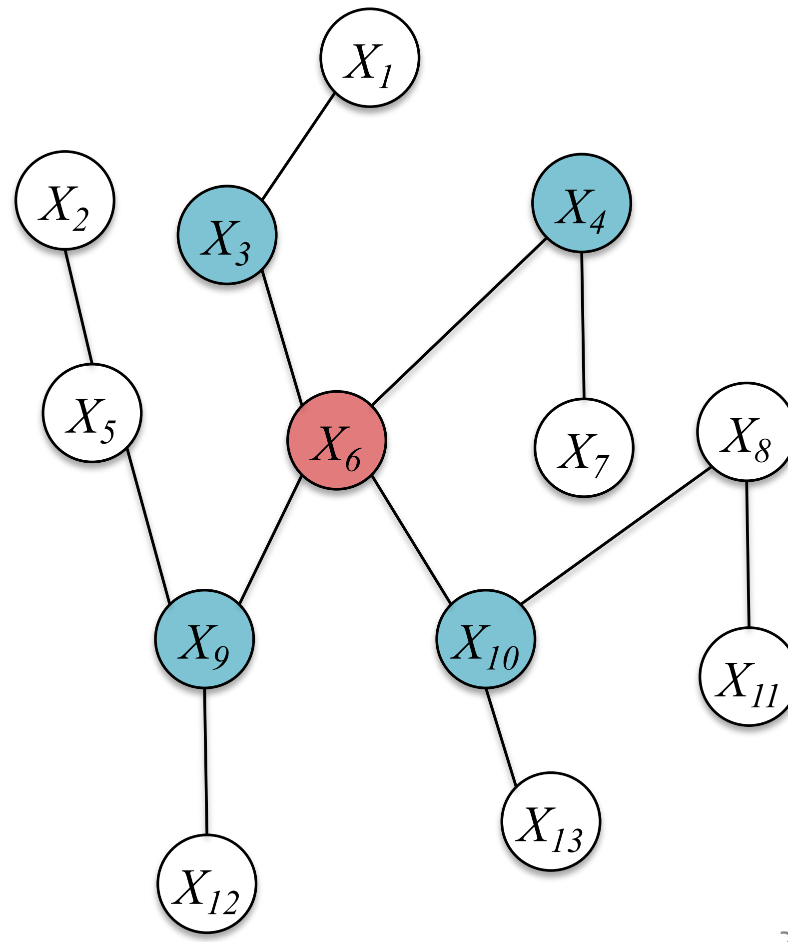


Markov Blanket (Undirected)

Def: the **Markov Blanket** of a node in an **undirected** graphical model is the set containing the node's neighbors.

Theorem: a node is **conditionally independent** of every other node in the graph given its **Markov blanket**

Example: The Markov Blanket of X_6 is $\{X_3, X_4, X_9, X_{10}\}$



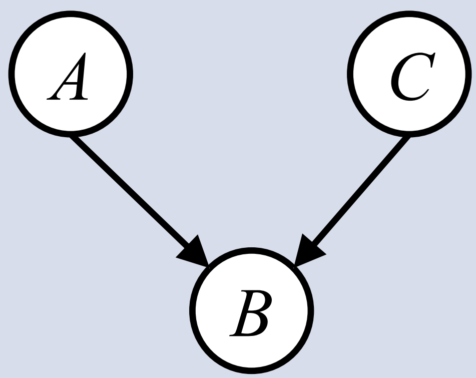
Undirected Graphical Models

Whiteboard

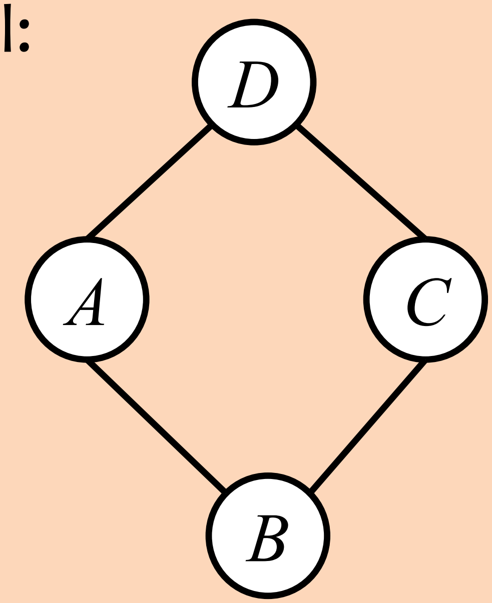
- Proof of independence by separation (simple case)
- Global Markov properties
- Hammersley-Clifford Theorem
- Local Markov properties
- Pairwise Markov properties
- Equivalent characterizations of UGMs

Non-equivalence of Directed / Undirected Graphical Models

There does **not** exist an **undirected** graphical model that can capture the conditional independence assumptions of this **directed** graphical model:



There does **not** exist a **directed** graphical model that can capture the conditional independence assumptions of this **undirected** graphical model:



Undirected Graphical Models

Whiteboard

- Alternate definition using maximal cliques
- Pairwise Markov Random Field (MRF)