CSCI E-82a
Probabilistic Programming and AI
Lecture 2
Markov Graphical Models

Steve Elston

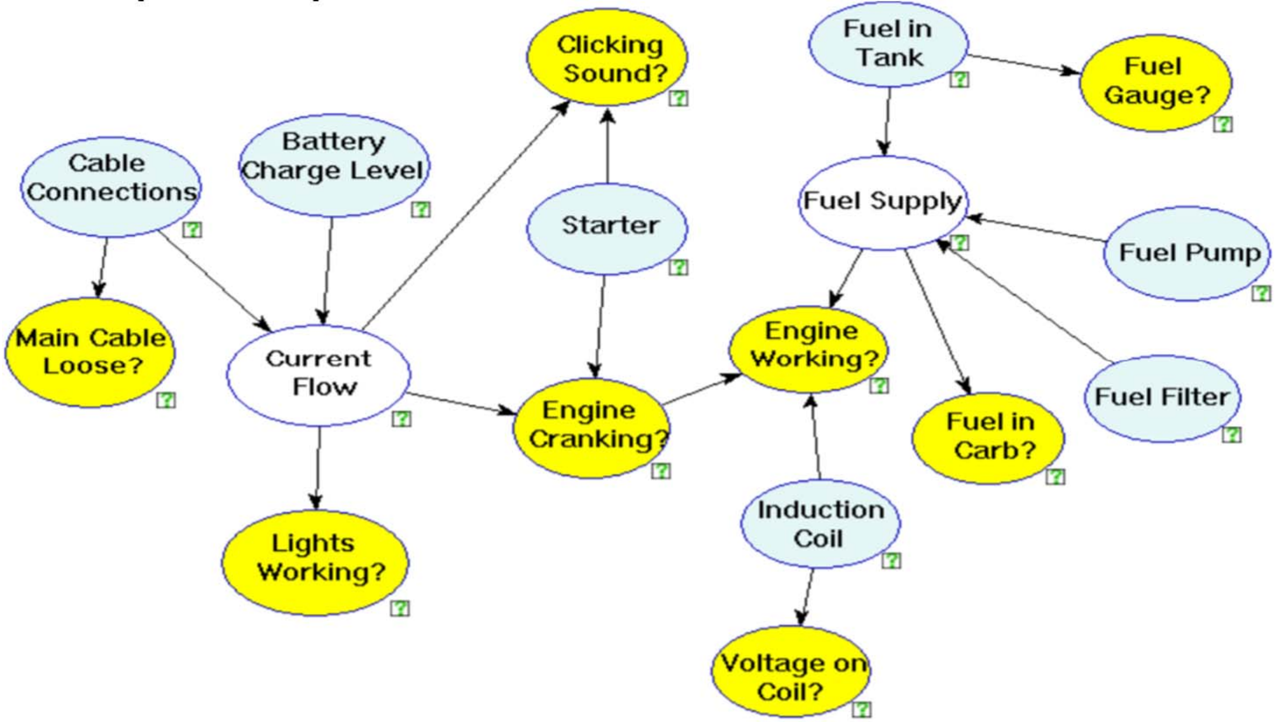
HARVARD
Extension School

Outline

- Why Markov Graphical Models?
- Properties of Markov Graphical Models
- Cliques of Graphical Models
- Potentials for Markov Graphical Models
- Independencies and the Hammerly Clifford Theorem
- Potentials and the Hammerly Clifford Theorem
- Independencies and Separation in MRFs
- Markov Blanket
- Pairwise Markov Property
- Transforming DAGs to MRFs and Moralization
- Independencies and Separation in MRFs
- DAGs vs MRFs
- Summary

Why Markov Graphical Models?

The previous lesson addressed **directed acyclic graphical** models (DAGs)



Attribution, Wojtek Przytula, et. al. 2002

Why Markov Graphical Models?

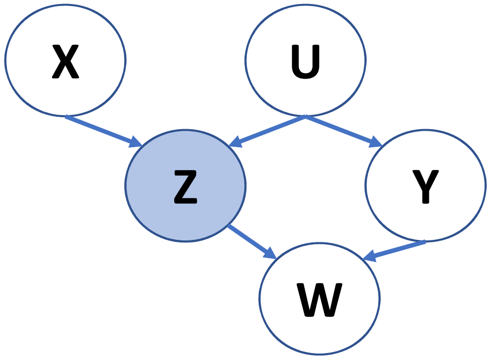
The previous lesson addressed **directed acyclic graphical** models (DAGs)

- DAGs **cannot represent certian independency structures:**
  - Non-directional dependency on neighbors
    - Spatial relationships
    - Social networks
    - Image data
    - Molecular structure
    - Many more…….
  - Cyclical structures
  - Etc…
- How can these independencies be represented?
- **Markov graphical model** or **Markov random field (MRF)** model

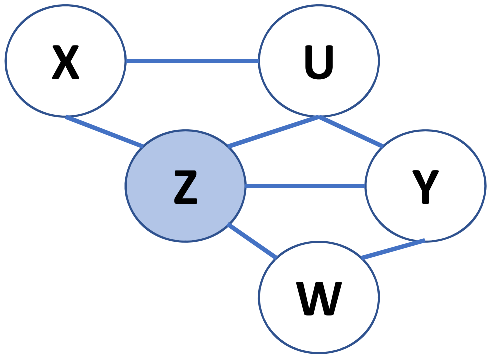
Why Markov Graphical Models?

Markov are related to DAGs

- DAGs have directed edges
- Markov graphical models have undirected edges
- DAGs can be transformed into equivalent Markov graph



DAG



Undirected Markov graph

Why Markov Graphical Models?

- Markov random field models arise in statistical physics
  - Solid state physics
  - e.g. Ising model of magnetism
- Markov random field models can represent complex **independency structure**
- But, the we pay a price in **computational complexity**
- If a DAG represents the independency structure, use it!
- Otherwise, we use a MRF, if computationally feasible

Why Markov Graphical Models?

Solve protein folding problem

http://www.yaroslavvb.com/papers/yanover-linear.pdf

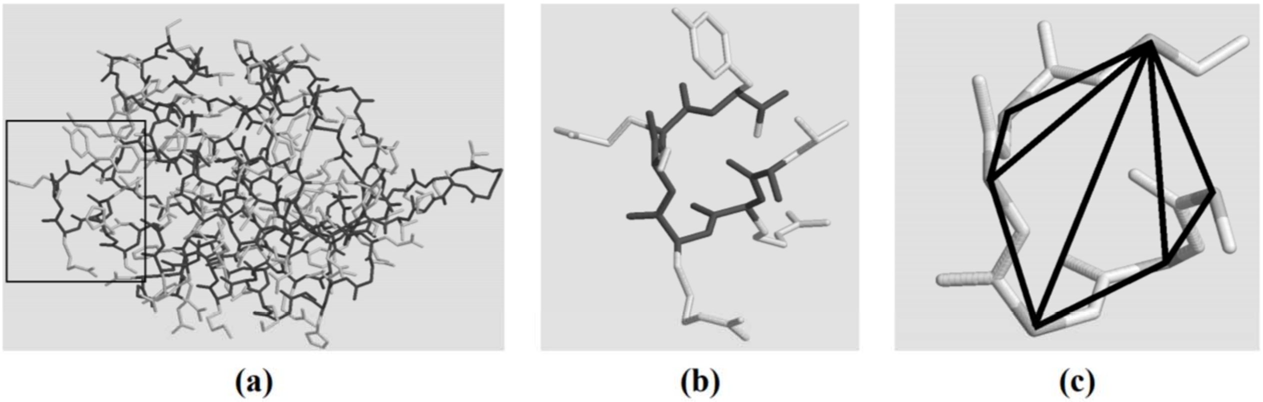
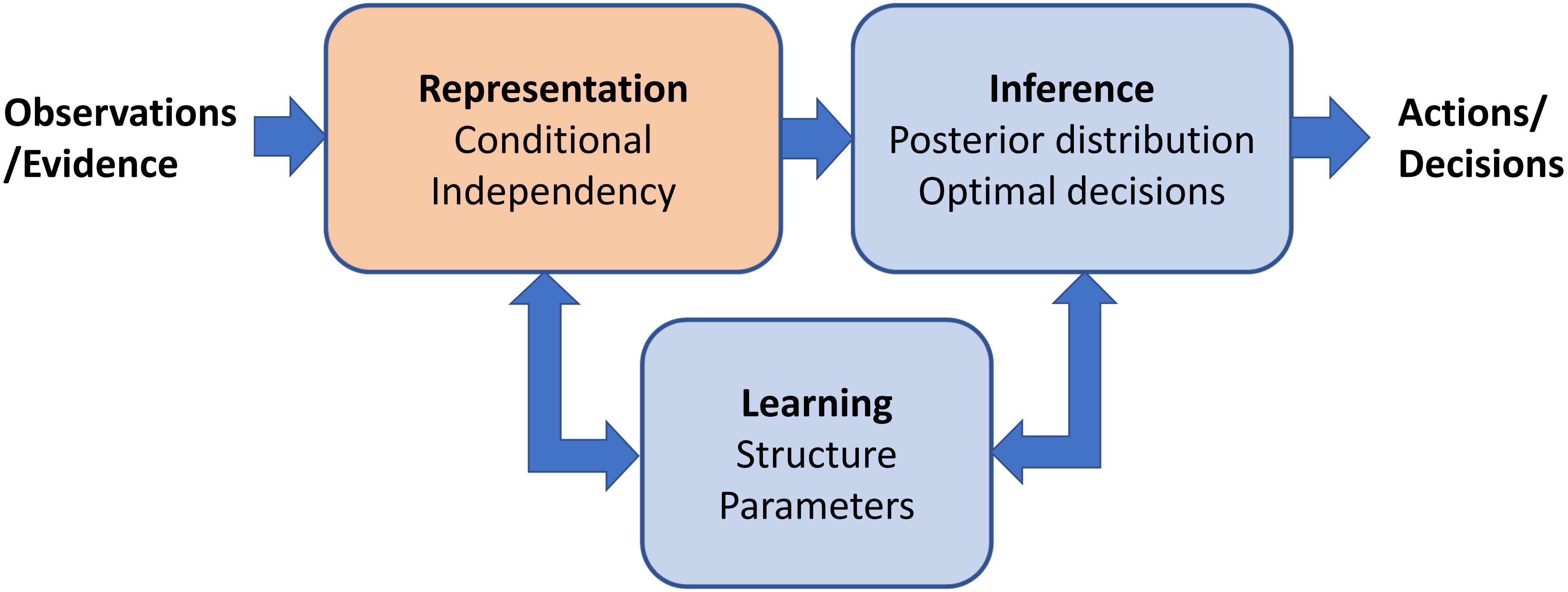


Figure 2: **(a)** Cow actin binding protein (PDB code 1pne). **(b)** A closer view of its 6 C-terminal residues. Given the protein backbone (black) and the amino acid sequence, side-chain prediction is the problem of predicting the native side-chain conformation (gray). **(c)** Problem representation as a graphical model for those C-terminal residues shown in (b) (nodes located at C^{α} atom positions, edges drawn in black).

Attribution; Yanover, Meltzer, Weiss 2006

Focus on representation with graphical models



Observations /Evidence

Representation
Conditional Independency

Inference
Posterior distribution
Optimal decisions

Actions/ Decisions

Learning
Structure
Parameters

Schematic of intelligent agent using directed graphical model

Properties of Markov Graphical Models

- Markov random field models have **undirected edges**
- Model distributions in MRFs using **potentials**
  - Recall that DAGs model distributions using **CPDs**
- Potentials are not distributions and must be normalized by a **partition function**
- There is a **potential** for **each clique of the graph**
- Potentials define the **strength of the interaction** between the nodes in a clique
  - For example, people who interact directly in a social network are more likely to influence each other (like a social clique)

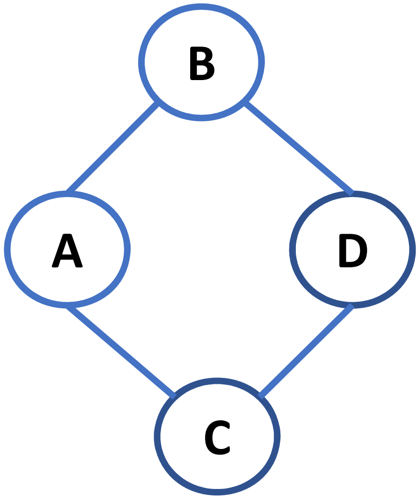
Cliques of Graphical Models

Definition: A **clique** is a subset of vertices of an undirected graph such that **every two distinct vertices in the clique are adjacent**

- The **subgraphs** of a clique **must be complete**
- **Independency structure** is determined by the cliques of the MRF
- A node can be in multiple cliques
- A clique can be as small as one node

Cliques of Graphical Models

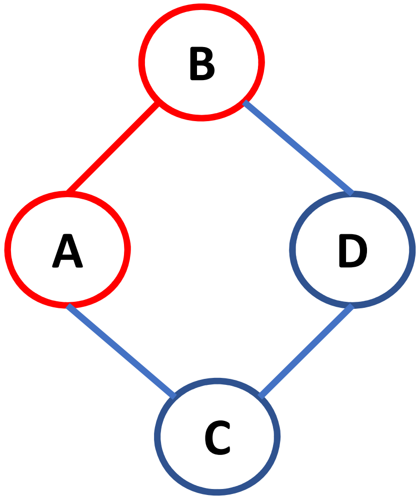
Definition: A **maximal clique** is a clique which cannot be enlarged without violating the clique property



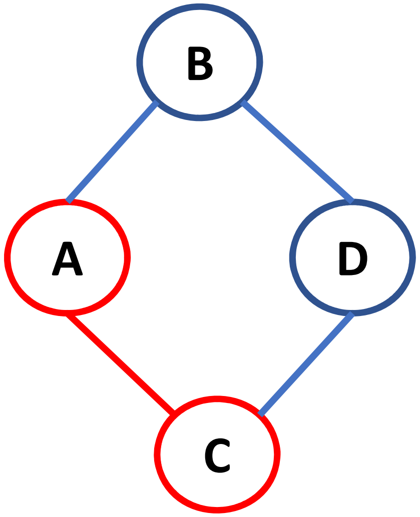
- Maximal Cliques required for inference algorithms
- Example: consider the undirected graph with cliques, (A,B), (A,C), (B,D), (C,D)
- Enlarging any of these cliques **violates the clique property**, since the vertices would not be adjacent
- Therefore, the cliques (A,B), (A,C), (B,D), (C,D) and **maximal cliques**

Cliques of Graphical Models - Example

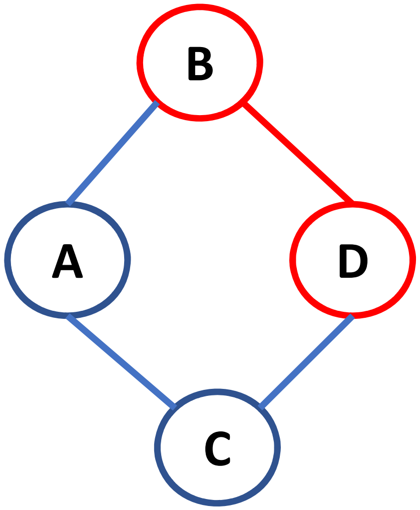
Example: undirected graph with 4 cliques defined



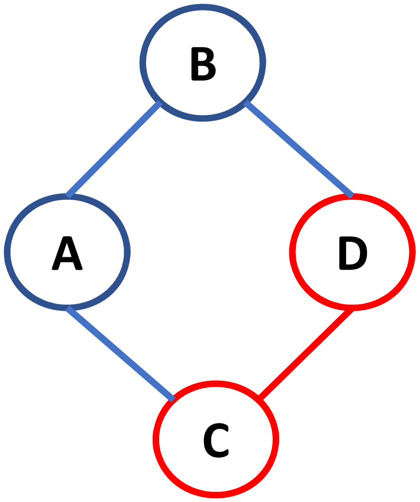
**Clique 1,
(A,B)**



**Clique 2,
(A,C)**



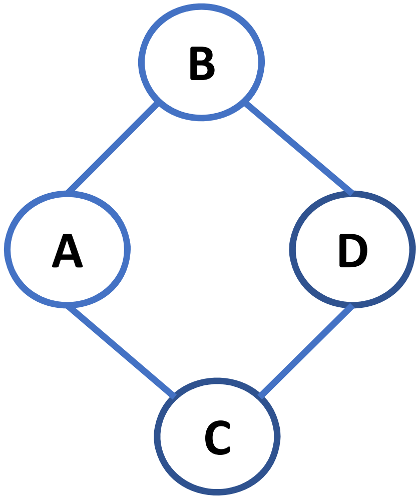
**Clique 3,
(B,D)**



**Clique 4,
(C,D)**

Potentials for Markov Graphical Models

Distribution for Markov random fields is modeled by **potentials**



- Each clique has a potential
- The **product of the potentials is also a potential**:

$$\tilde{p}(A, B, C, D) = \phi(A, B)\phi(A, C)\phi(B, D)\phi(C, D)$$

- The product of the potentials can be **transformed to a distribution by a normalization**:

$$p(A, B, C, D) = \frac{1}{Z}\tilde{p}(A, B, C, D)$$

- Where the normalization is the **partition function:**

$$Z = \sum_{A,B,C,D} \tilde{p}(A, B, C, D)$$

Potentials for Markov Graphical Models

How to formulate a **distribution given the potentials** for complex graphs?

- The general formulation for a multivariate distribution is

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c)$$

- Where, c is a clique in the set of cliques
- The partition function is given by:

$$Z = \sum_{x_1, \ldots, x_n} \prod_{c \in C} \phi_c(x_c)$$

Potentials for Markov Graphical Models

Computing the partition function presents a significant problem

$$Z = \sum_{x_1,\ldots,x_n} \prod_{c \in C} \phi_c(x_c)$$

- Partition function has high computational complexity
  - Compute a product for each data sample
  - Sum the products over the data samples
- Computation complexity makes exact solution of many MRF problems impractically difficult

Potentials for Markov Graphical Models

How does all this relate to Bayesian networks?

- Recall, we can express the distribution for a Bayesian network using global semantics:

$$P(X) = \prod_{i=1:d} P(X_i | \{parents(X_i)\})$$

- The potentials are the conditional probability distributions (CPDs)
- But, what happened to the partition function?
- For Bayesian networks Z = 1.0

Independencies and the Hammerly Clifford Theorem

How can we **model independencies** in undirected graphical models?

- The **Hammerly-Clifford theorem** provides a tool to map the **conditional independence properties** of an undirected graph G
- Further, the Hammerly-Clifford theorem gives a practical way to formulate potentials

Independendencies and the Hammerly Clifford Theorem

Hammerly-Clifford theorem: Let p(x) be a strictly positive distribution and let G be an undirected graph, the **conditional independence properties** of p(x) are satisfied if and only if the distribution can be represented as a product of factors, one factor representing each maximal clique, c, of G:

$$p(x \mid \Theta) = \frac{1}{Z(\theta)} \prod_{c \in G} \psi_c(x_c \mid \theta_c)$$

Where,

$$Z(\theta) = \sum_{x} \prod_{c \in G} \psi_c(x_c \mid \theta_c)$$

And, $Z(\theta)$ is the partition function that ensures $p(x \mid \Theta)$ is in the range $\{0, 1\}$

Potentials and the Hammerly Clifford Theorem

The Hammerly Clifford Theorem provides a practical way to formulate potentials

- Use a **Gibbs Distribution**:

$$p(x \mid \theta) = \frac{1}{Z(\theta)} exp\left(- \sum_{c} E(x_c \mid \theta_c) \right)$$

- Where, $E(x_c) = Energy\, of\; clique\; c$
- The potential for a clique is then:

$$\phi(x_c \mid \theta_c) = exp\left(- E(x_c \mid \theta_c) \right)$$

- Given the minus sign, **the lower the energy of the state of a clique, c, the higher the probability**

Independencies and Separation in MRFs

How can we **model independencies or separation** in undirected graphical models?

- **Definition:** For a graph G, **disjoint** subsets A and B are separated by subset S if every path from A to B passes through S then S **separates** A and B. Or, if $S = \emptyset$, then no path exists from A to B, and A and B are **separated**
- **Definition:** For a graph G with disjoint sets A, B and S, where S separates A and B, then $A \perp B \mid S$, known as the **global Markov property**

Independencies and Separation in MRFs

How can we **model independencies or separation** in undirected graphical models?

- **Definition:** Given subsets X, Y and Z, X and Y are conditionally independent or **D-separated** conditioned on the subset Z if they are separated on the moralized graph
- **Definition:** A graph G is a **dependency map** or **D-map** of a distribution P if the graph contains every conditional independence in P. We can represent this relationship as:

$$(X \perp Y \mid Z_G) \Leftarrow (X \perp Y \mid Z_P)$$

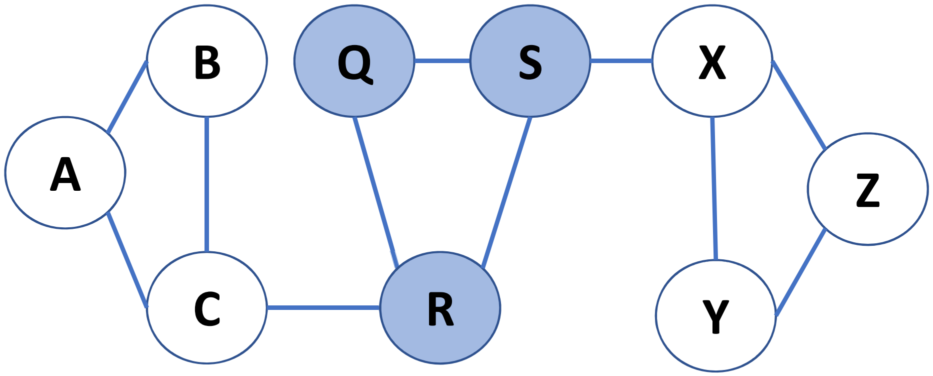
Independencies and Separation in MRFs

There are two significant properties of independence maps in undirected graphs that are **soundness** and **completeness**

- **Theorem:** For any graph G that factorizes a distribution P then $I(G) \subseteq I(P)$. This relationship is known as the **soundness** property
- **Claim:** For any graph G, with subsets X, Y and Z, that factorizes a distribution P, if $(X \perp Y \mid Z) \subseteq I(P)$ then d-$sep_G(X; Y \mid Z)$. This relationship is known as the **completeness property**

Independencies and Separation in MRFs - Example

Example of **modeling independencies or separation** in undirected graphical models



There are **5 maximal cliques**

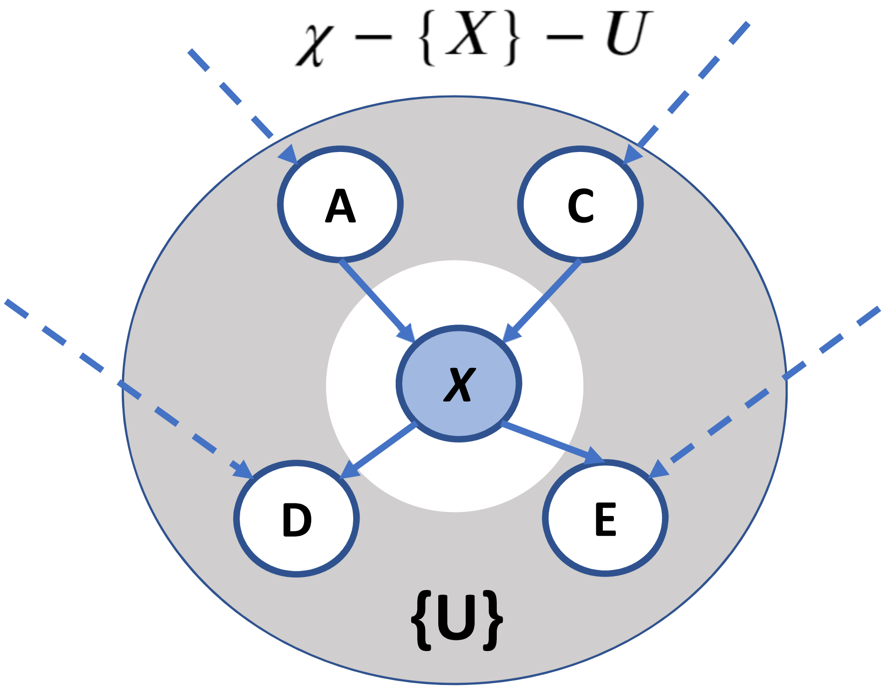
{A,B,C}
{C,R}
{Q,R,S}
{S,X}
{X,Y,Z}

Cliques {A,B,C} and {X,Y,Z} are separated by clique {Q,R,S}

Or, $\{A, B, C\} \perp \{X, Y, Z\} \mid \{Q, R, S\}$

Markov Blanket

For a DAG, any node is conditionally independent or all others given its **Markov Blanket**



Definition: A subset *U* is a **Markov blanket** of *X* in the set of nodes χ of the graph G if $X \notin U$ and if *U* is a minimal set of nodes such that:

$$(X \perp \chi - \{X\} - U \mid U) \in I(P)$$

- Where $\chi - \{X\} - U$ is the set of nodes not in *X* or *U*
- This definition is a result of the D-separation property for MRFs

Pairwise Markov Property

The **pairwise Markov property** connects the local and global Markov properties of MRFs

Definition: Two nodes are conditionally independent given the other nodes in the graph if there is no direct edge between them. This property is the **pairwise Markov property**

- The pairwise Markov property relates to the **global Markov property** and the **local Markov property** in a somewhat circular fashion:

$$Global \rightarrow Local \rightarrow Pairwise \rightarrow Global$$

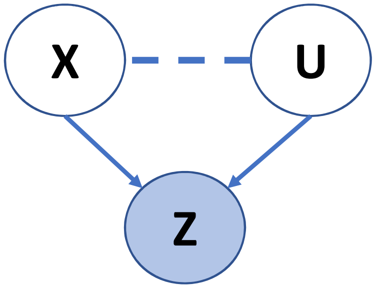
Transforming DAGs to MRFs and Moralization

How is the structure of a DAG related to the structure of MRF

- The process of **moralization** is the key step in transforming a DAG to an MRF
- **Definition:** An **immorality** in a directed graph G occurs where either; a) there is a directed edge between X and Y, or b) X and Y are both parents of the same note Z
- **Definition:** A **moral graph**, M(G), of a BN structure, G is the **undirected graph** over X that contains an undirected edge between X and Y if; a) there is a directed edge between X and Y, or b) X and Y are both parents of the same note Z

Transforming DAGs to MRFs and Moralization - Example

Example of an immorality



- Consider a DAG with a **v-structure or collider**
- With Z not observed the independency can be expressed:

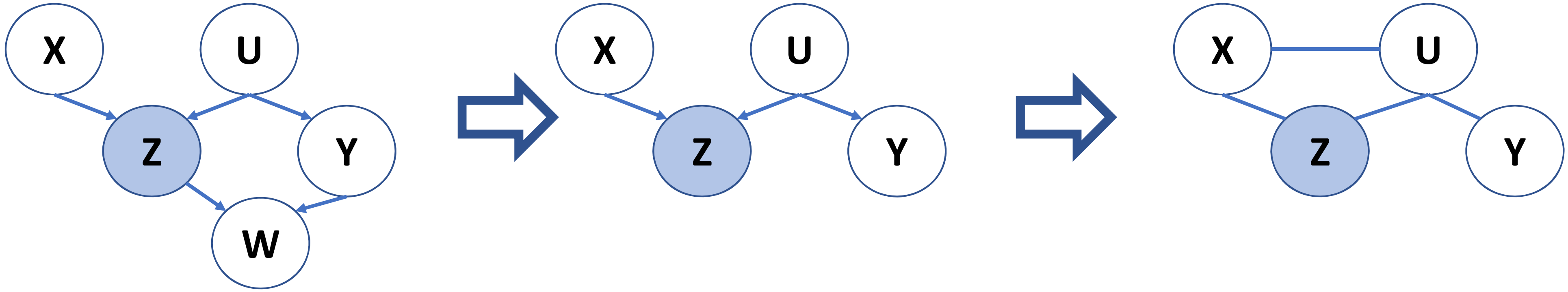
$$P(X, U \mid Z) = P(X \mid Z)\ P(U \mid Z)$$

- The path from X to U is blocked by Z
- Therefore, this relationship is an **immorality** since X and U are parents of Z
- We **moralize** the graph by **marrying** X and U with an undirected edge

Transforming DAGs to MRFs and Moralization - Example

Example of transforming a DAG to a MRF

- Start with the original DAG
- Not observing Z blocks a path to W and U and X are independent
- Therefore, the **ancestral graph** does not contain W
- Moralized undirected graph



Original DAG

Ancestral graph

Moralized undirected graph

Transforming DAGs to MRFs and Moralization

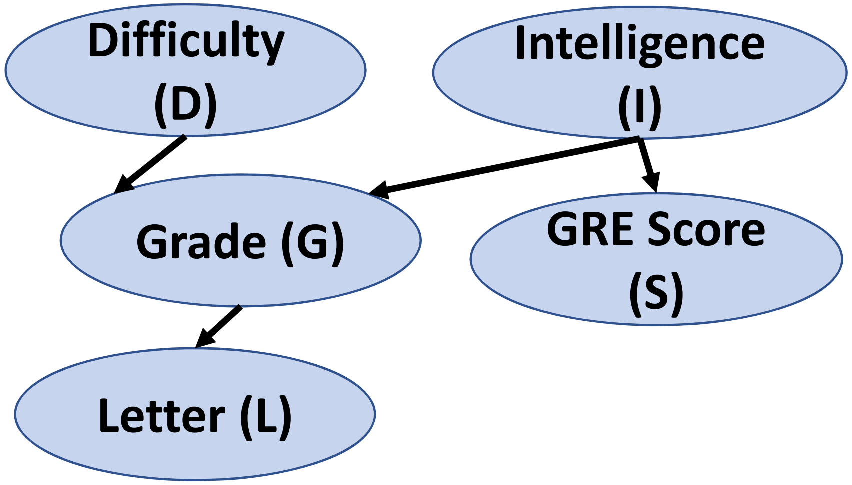
How is the structure of a DAG related to the structure of MRF

- We can relate the I-map between a DAG and a MRF though a corollary
- **Corollary:** Given a distribution P_B such that B is a parameterization on a graph G, then M(G) is an I-map for P_B
- However, all of this **does not mean** that the independency structure of the DAG and resulting MRF will be the same

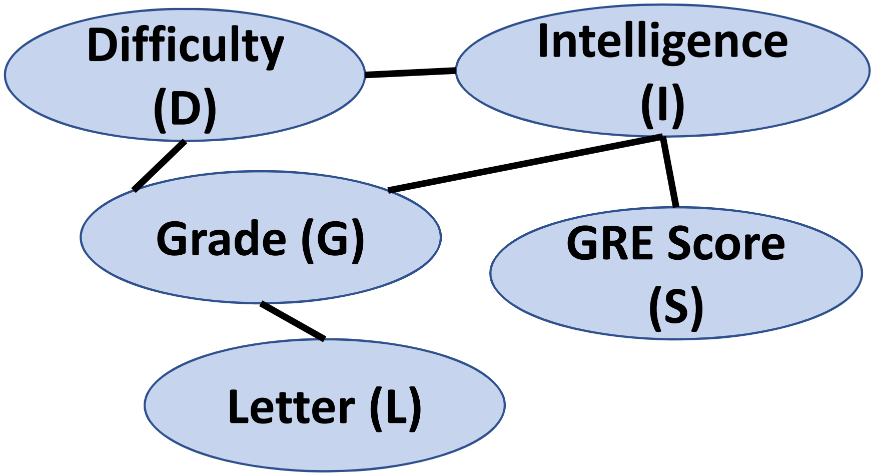
Independencies and Separation in MRFs - Example

Example of **modeling independencies or separation** in undirected graphical models

- Start with DAG
- Transform to undirected moralized graph



DAG

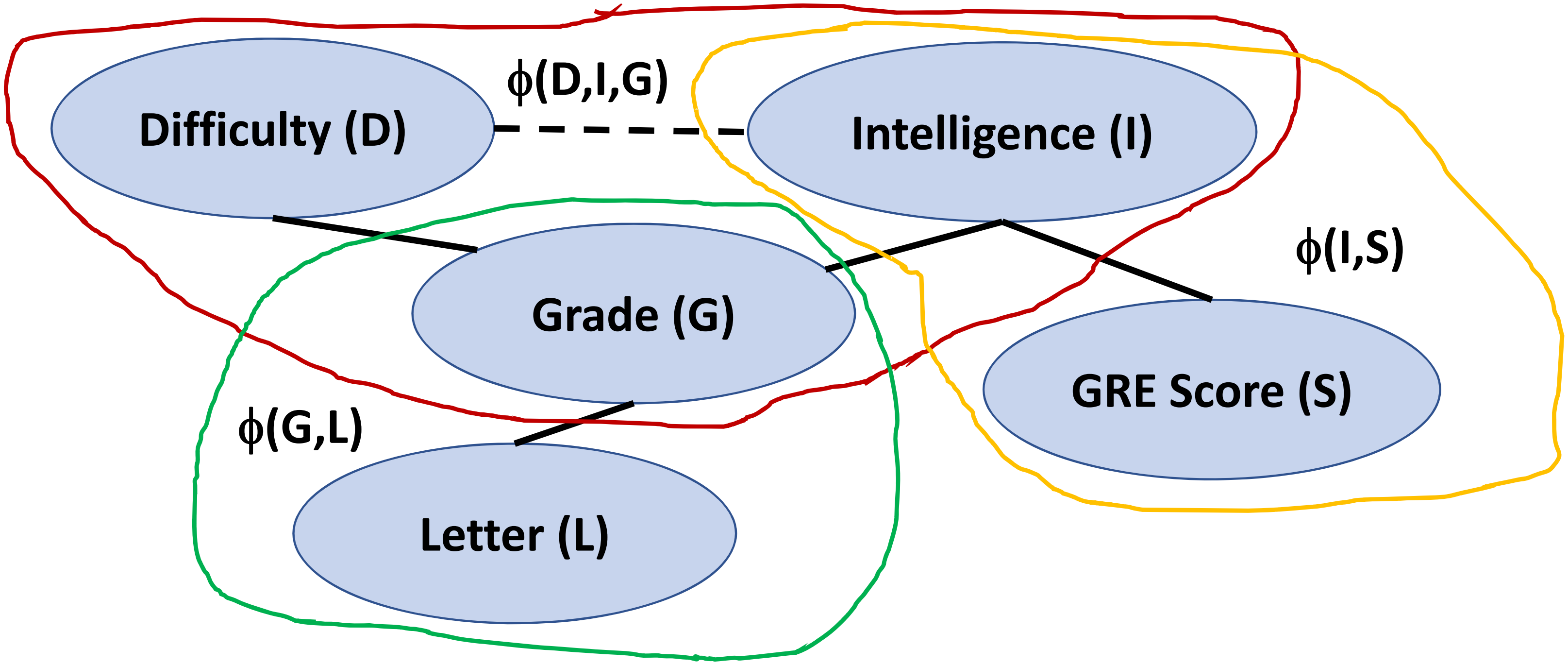


Moralized MN

Independencies and Separation in MRFs - Example

Example of **modeling independencies or separation** in undirected graphical models

- Define maximal cliques on moralized undirected graph
- Each clique has a potential



Independencies and Separation in MRFs - Example

Example of **modeling independencies or separation** in undirected graphical models

- Now, factorize the unconditional distribution into potentials

$$
\begin{aligned}
P(I, D, G, S, L) &= \frac{1}{Z}\phi(D, I, G)\ \phi(G, L)\ \phi(I, S) \\
&= \frac{1}{Z} exp\{-\mathbb{E}(D, I, G) - \mathbb{E}(G, L) - \mathbb{E}(I, S)\}
\end{aligned}
$$

And the partition function is:

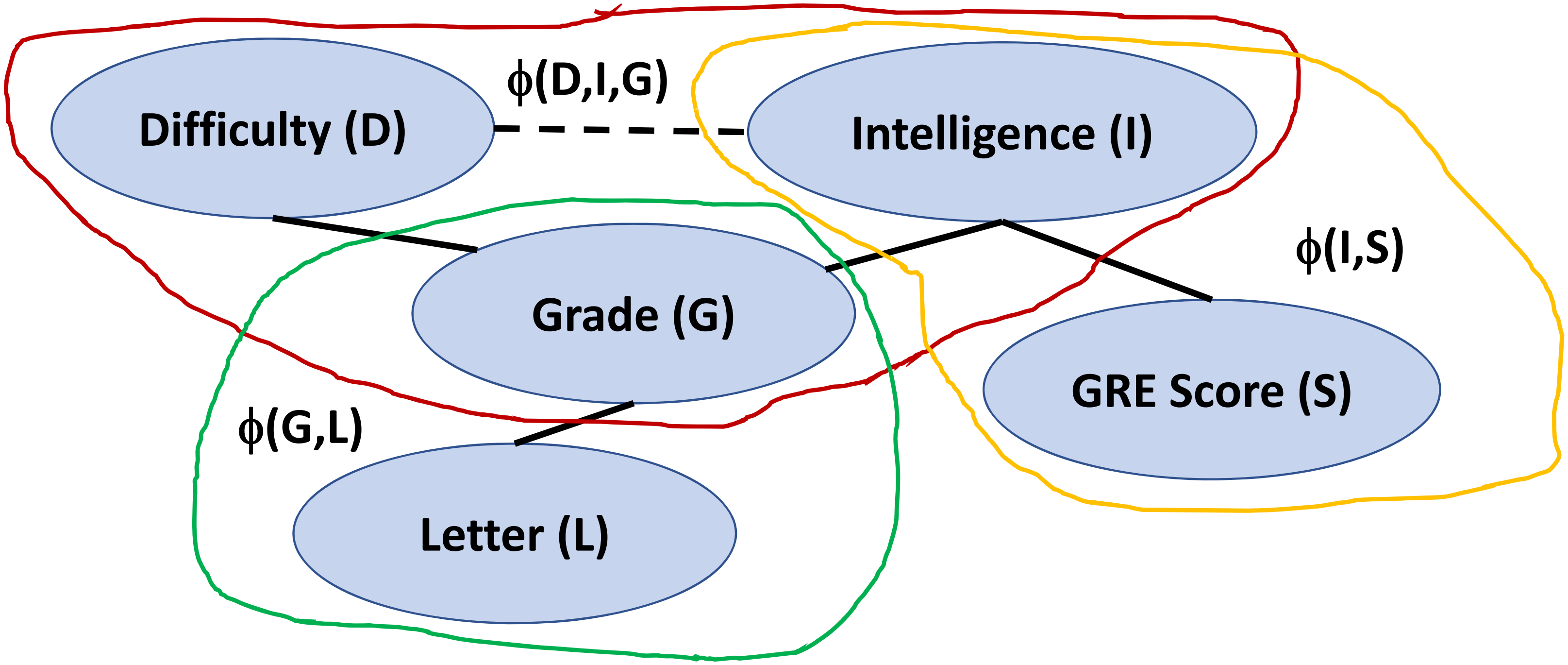
$$
Z = \sum_{I,D,G,S,L} \phi(D, I, G)\ \phi(G, L)\ \phi(I, S)
$$

The distribution is modeled at the product of potentials on the undirected graph

DAGs vs MRFs - Example

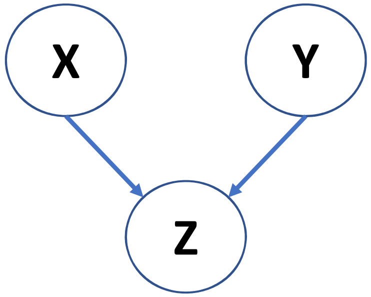
Example of **modeling independencies or separation** in undirected graphical models

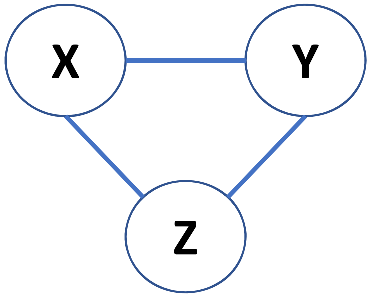
- Wait! Have we **lost the independence map** of the DAG?
- Yes**, the potential of the maximal cliques have a different map!**



DAGs vs MRFs

Directed acyclic graphs and Markov networks **cannot always represent the same independency structures**

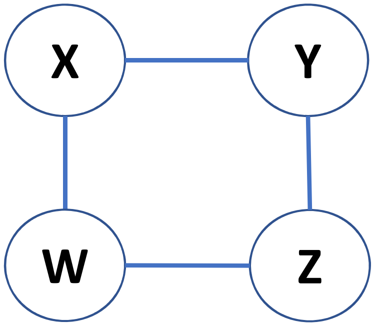


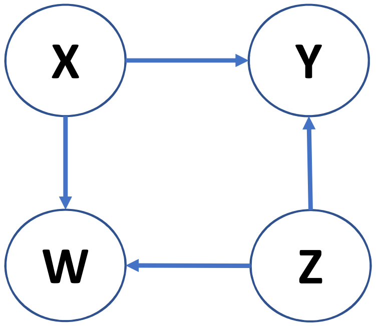


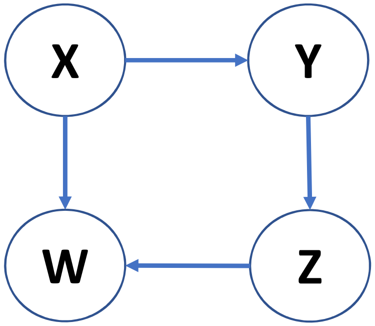
- Why is this the case?
- Consider a DAG with a **v-structure or collider**
- **X and Y are unconditionally independent:** $X \perp Y$.
- Now consider the moralized Markov network
- There is **no independency!**

DAGs vs MRFs

Directed acyclic graphs and Markov networks **cannot always represent the same independency structures**







- Consider the Markov network with 4 maximal cliques, and potentials:

$$\phi(W,X)\ \phi(X,Y)\ \phi(Y,Z)\ \phi(Z,W)$$

Independencies are:

$$\phi(X,Y) \perp \phi(Z,W) \mid \{\phi(W,X), \phi(Y,Z)\}$$

$$\phi(W,X) \perp \phi(Y,Z) \mid \{\phi(X,Y), \phi(Z,W)\}$$

- Multiple DAGs are possible on this skeleton, eg.
- But, **no DAG can represent the independencies!**

DAGs vs MRFs

Directed acyclic graphs and Markov networks **cannot always represent the same independency structures**

There are actually **4 possible cases of independency maps**:

- Representable by a DAG, but not a MRF
- Representable by a MRF, but not a DAG
- Representable by both a DAG and MRF
- Not representable by either a DAG or MRF – fortunately, rate in practice

Vocabulary Summary

A **clique** is a subset of vertices of an undirected graph such that **every two distinct vertices in the clique are adjacent**

A **maximal clique** is a clique which cannot be enlarged without violating the clique property

For a graph G, **disjoint** subsets A and B are separated by subset S if every path from A to B passes through S then S **separates** A and B. Or, if $S = \emptyset$, then no path exists from A to B, and A and B are **separated**

For a graph G with disjoint sets A, B and S, where S separates A and B, then $A \perp B \mid S$, known as the **global Markov property**

Given subsets X, Y and Z, X and Y are conditionally independent or **D-separated** conditioned on the subset Z if they are separated on the moralized graph

A graph G is a **dependency map** or **D-map** of a distribution P if the graph contains every conditional independence in P.

A subset *U* is a **Markov blanket** of *X* in the set of nodes χ of the graph G if $X \notin U$ and if *U* is a minimal set of nodes such that: $(X \perp \chi - \{X\} - U \mid U) \in I(P)$

Two nodes are conditionally independent given the other nodes in the graph if there is no direct edge between them. This property is the **pairwise Markov property**

Moralization is a key step in turning DAG to an MRF

An **immorality** in a directed graph G occurs where either; a) there is a directed edge between X and Y, or b) X and Y are both parents of the same note Z

A **moral graph**, M(G), of a BN structure, G is the **undirected graph** over X that contains an undirected edge between X and Y if; a) there is a directed edge between X and Y, or b) X and Y are both parents of the same note Z