INFO510 Bayesian Modelling and Inference

Lecture 9 – Markov Random Fields

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

September 29, 2025

INFO510 Bayesian Modelling and Inference

Lecture 9 – Markov Random Fields

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

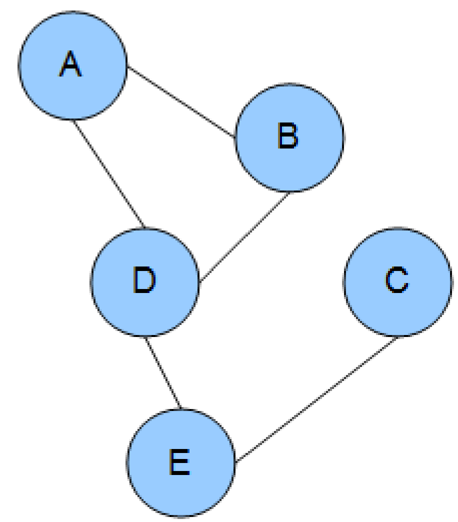
September 29, 2025

Markov Random Field (MRF)

Graphical model used to represent the joint distribution of a set of random variables

It's a way of simplifying complex relationships between items by focusing on local dependencies, meaning how closely related things directly influence each other.

It is an undirected network

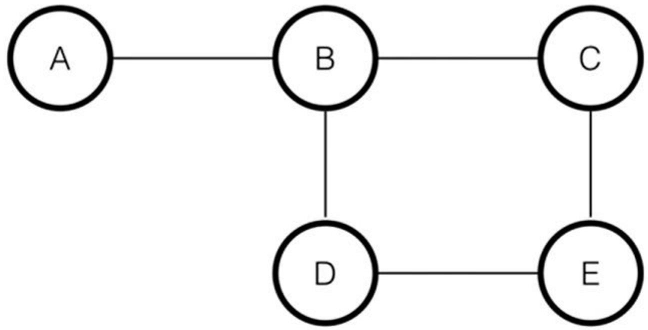


Two nodes are connected if they are not independent conditional on all other nodes

More importantly, two nodes are NOT connected if they are independent conditioned on all nodes

A node separates two nodes if it is on path from one node to another

Undirected Graphical Models



$$P(A, B, C, D, E) \propto \phi(A, B)\phi(B, C)\phi(B, D)\phi(C, E)\phi(D, E)$$

$$P(X) = \frac{1}{Z} \prod_{c \in \text{cliques}(G)} \phi_C(x_c)$$

Clique (graph theory)

21 languages

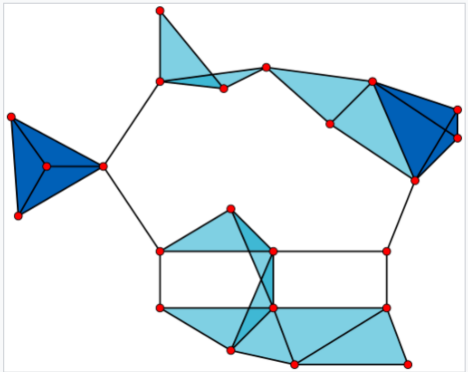
Article Talk Read Edit View history Tools

Contents hide

From Wikipedia, the free encyclopedia

For other uses, see Clique (disambiguation).

In graph theory, a **clique** (/ˈkliːk/ or /ˈklɪk/) is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent. That is, a clique of a graph G is an induced subgraph of G that is complete. Cliques are one of the basic concepts of graph theory and are used in many other mathematical problems and constructions on graphs. Cliques have also been studied in computer science: the task of finding whether there is a clique of a given size in a graph (the clique problem) is NP-complete, but despite this hardness result, many algorithms for finding cliques have been studied.



A graph with

- 23 × 1-vertex cliques (the vertices),
- 42 × 2-vertex cliques (the edges),
- 19 × 3-vertex cliques (light and dark blue

Although the study of complete subgraphs goes

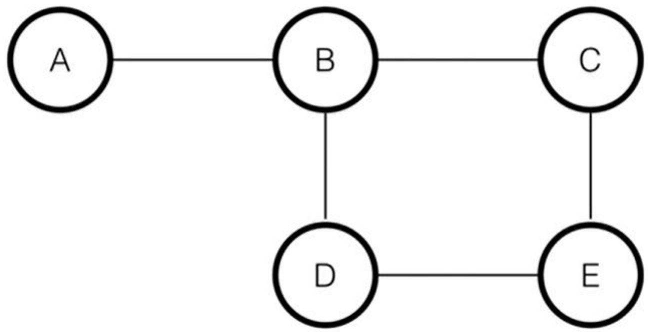
Appearance hide

Text
Small
Standard
Large

Width
Standard
Wide

Color (beta)
Automatic
Light
Dark

Undirected Graphical Models

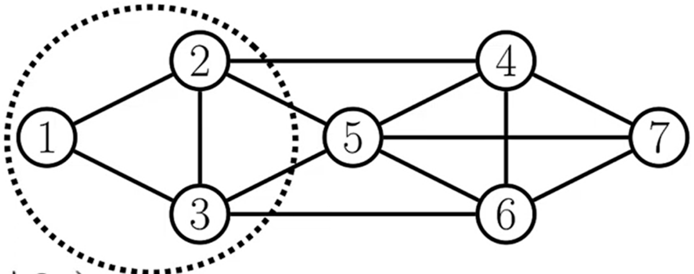


$$P(A, B, C, D, E) \propto \phi(A, B)\phi(B, C)\phi(B, D)\phi(C, E)\phi(D, E)$$

$$P(X) = \frac{1}{Z} \prod_{c \in \text{cliques}(G)} \phi_C(x_c)$$

Potential function

Markov Random Field (MRF)



$$\psi_c(\mathbf{y}_c|\boldsymbol{\theta}_c)$$

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{y}_c|\boldsymbol{\theta}_c)$$

$$Z(\boldsymbol{\theta}) \triangleq \sum_{\mathbf{y}} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{y}_c|\boldsymbol{\theta}_c)$$

Clique potential for every clique in the graph

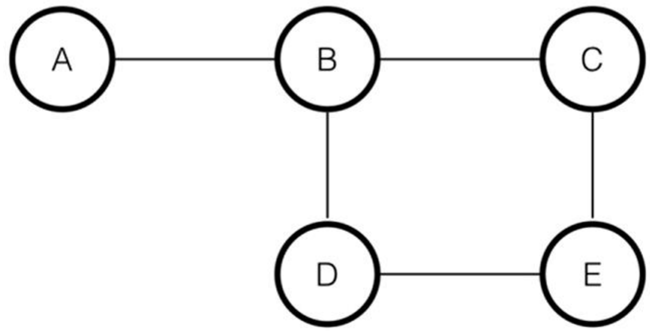
Clique is a set of variables that are all connected to one another in the graph

A potential function is defined for each clique

It takes the values of the variables in the clique and gives a number

Z = normalising constant

Undirected Graphical Models

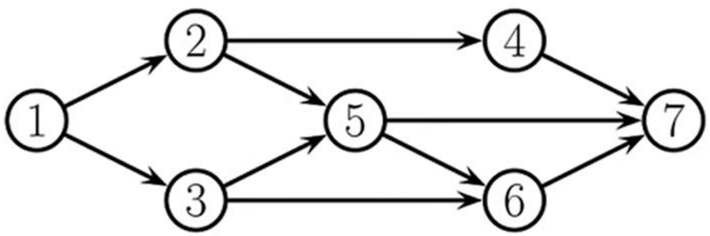


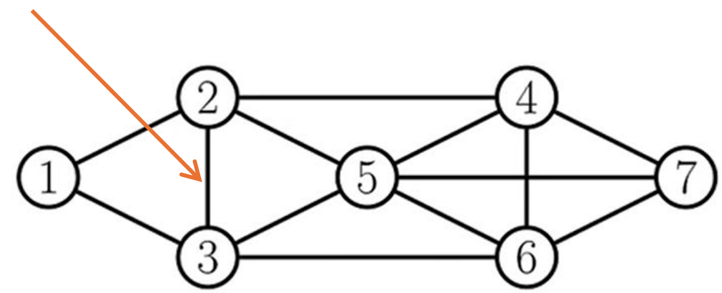
$$P(A, B, C, D, E) \propto \phi(A, B)\phi(B, C)\phi(B, D)\phi(C, E)\phi(D, E)$$

$$P(X) = \frac{1}{Z} \prod_{c \in \text{cliques}(G)} \phi_C(x_c)$$

potential functions

Converting Bayesian Networks (DAG) to Markov Random Field





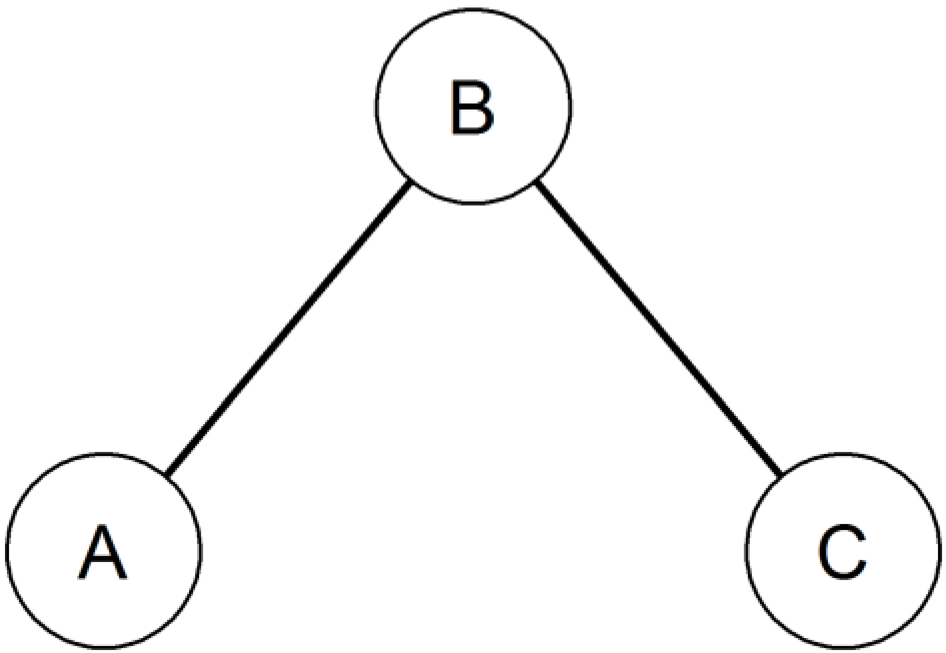
P(1)
P(2|1)
P(3|1)
P(5|2,3)

ψ1(1)
ψ2(1,2)
ψ3(1,3)
ψ5(2,3,5)

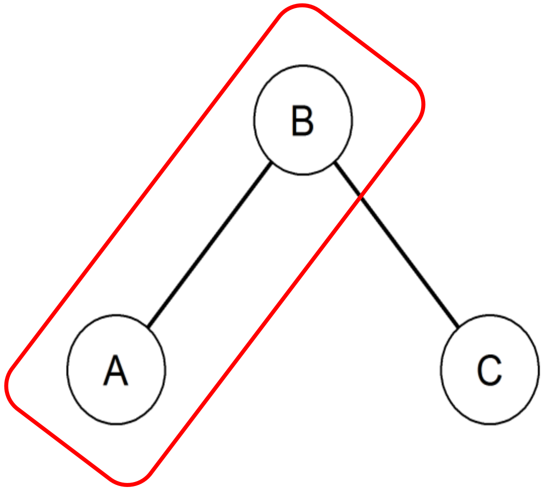
P(1:7) = P(1)*P(2|1)*P(3|1)*P(5|2,3).........P(7|4,5,6)

Convert a Bayesian network into MRF by moralizing

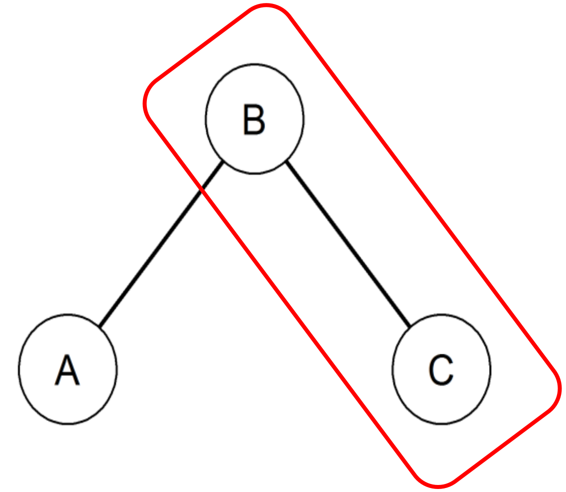
Conditional independence relationships



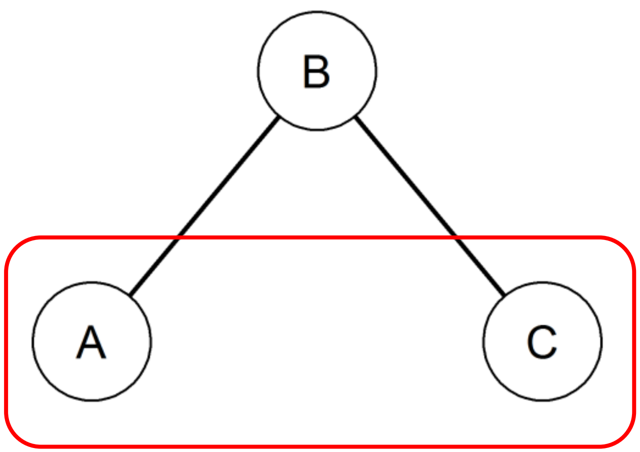
- B separates A and C
- $A \perp\!\!\!\perp C \mid B$



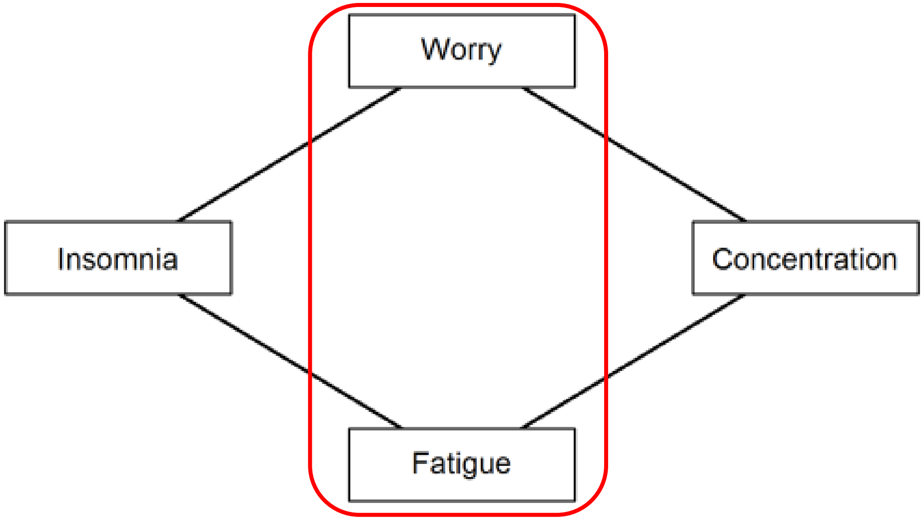
A and B are conditionally dependent, when controlled for C



B and C are conditionally dependent, when controlled for A



A and C are conditionally independent when controlled for B

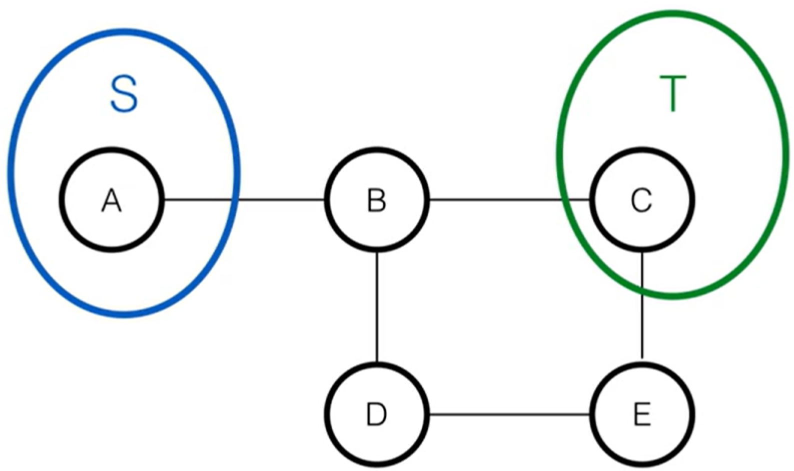


- Worrying and fatigue separate Insomnia and Concentration

Some relationships are better when modelled on undirected graphs, than on directed graphs

Markov Random Fields

- Any two subsets S and T of variables are conditionally independent given a **separating subset**
  - All paths between S and T must travel through the separating subset

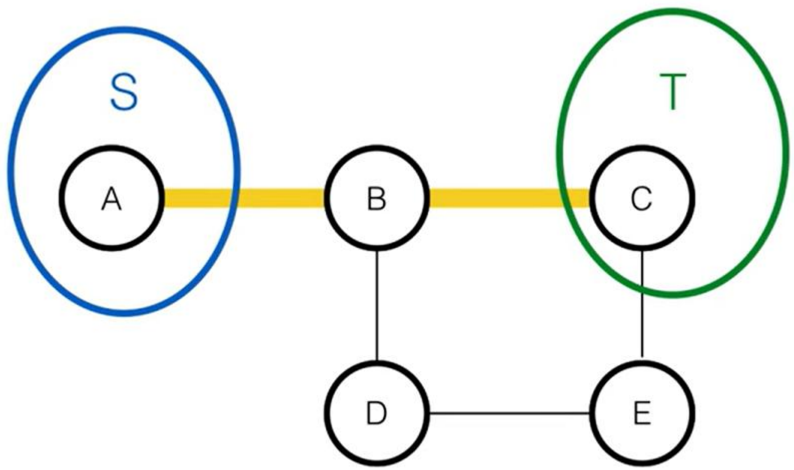


Markov Random Fields

- Any two subsets S and T of variables are conditionally independent given a **separating subset**
  - All paths between S and T must travel through the separating subset

paths:

A-B-C



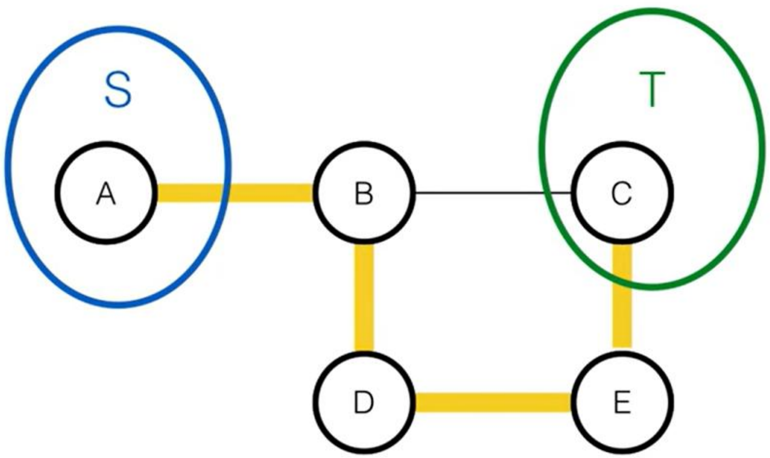
Markov Random Fields

- Any two subsets S and T of variables are conditionally independent given a **separating subset**
  - All paths between S and T must travel through the separating subset

paths:

A-B-C

A-B-D-E-C



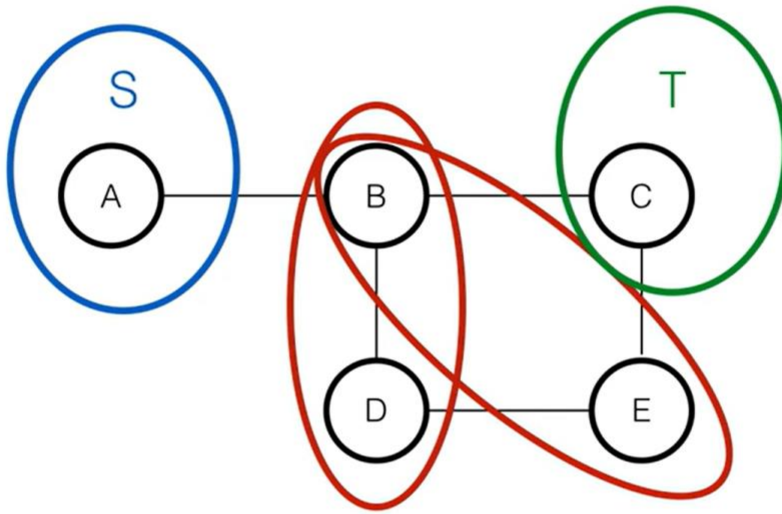
Markov Random Fields

- Any two subsets S and T of variables are conditionally independent given a **separating subset**
    - All paths between S and T must travel through the separating subset

paths:

A-B-C

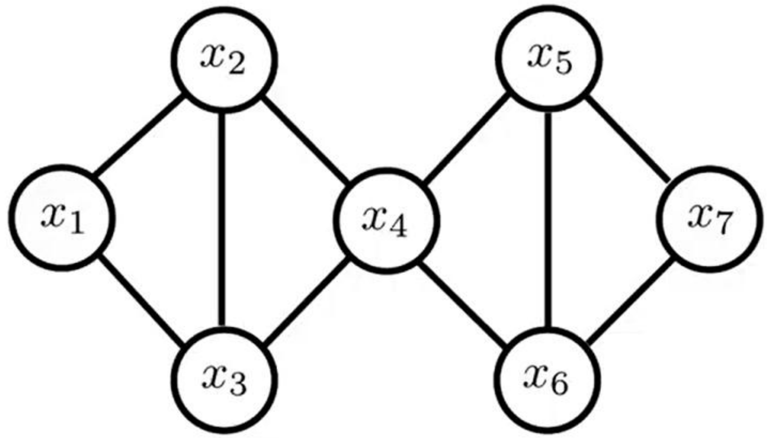
A-B-D-E-C



separating subsets

{B,D}, {B,E}, {B,D,E}

Local Markov Property – Example



- $p(x_4 \mid x_1, x_2, x_3, x_4, x_5, x_6, x_7) = p(x_4 \mid x_2, x_3, x_5, x_6)$
- In other words $x_4 \perp\!\!\!\perp \{x_1, x_7\} \mid \{x_2, x_3, x_5, x_6\}$
- Similarly, other independence relationships can be read off the graph

Key Concept: The Markov Property

The future state of a system depends **only** on its current state and not on any of the past states

Allows us to simplify the analysis of complex systems by reducing the number of variables we need to consider in our analysis

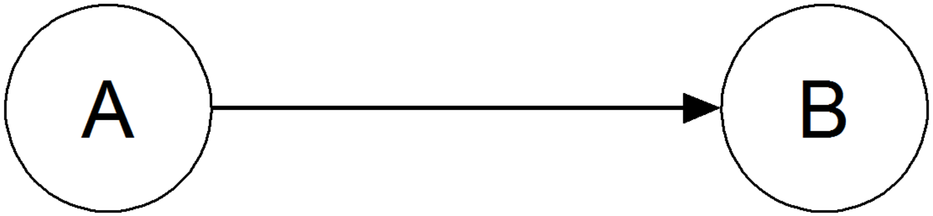
Intuitively meaningful - our current state already captures the information of the past states.

A Markov random field extends this property to two or more dimensions or to random variables defined for an interconnected network of items.

Interpreting a Markov Random Field

The edges in a MRF can be interpreted in several ways:

- Predictive effects
- Pairwise interactions
- Genuine symmetric relationships between nodes
  - Ising Model

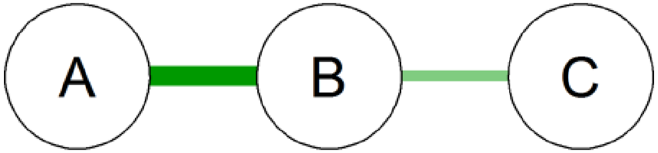


Predictive Effects

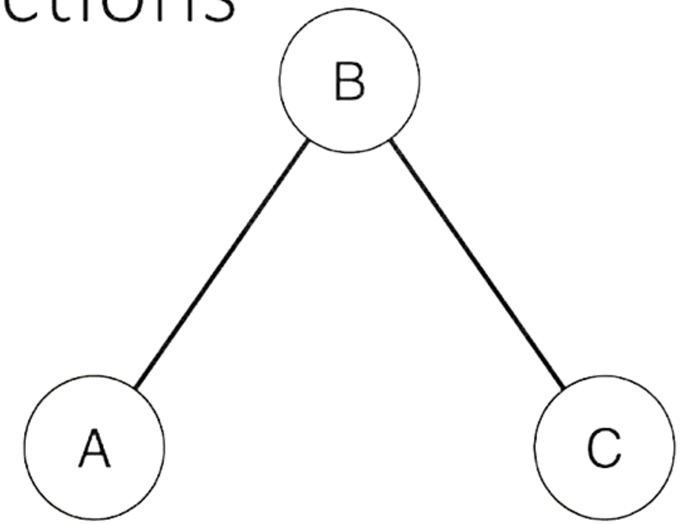
A MRF allows you to read predictive effects, close connection with regression models

A predicts B and B predicts A

A predicts B better than C predicts B
The relationship between A and C is mediated by B
It can also give you insights in multicollinearity –
What predicts the predictors?



Pairwise interactions



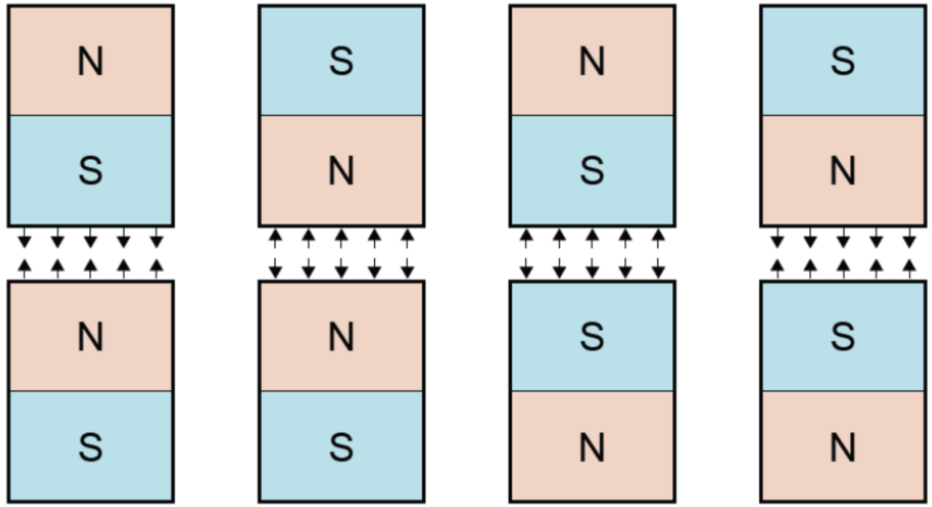
- If A and B interact, and B and C interact, then A and C are expected to be correlated

Connection to causal models

- The MRF model:
  - Concentration − Fatigue − Insomnia

- Is equivalent to three causal structures:
  1. Concentration → Fatigue → Insomnia
  2. Concentration ← Fatigue → Insomnia
  3. Concentration ← Fatigue ← Insomnia

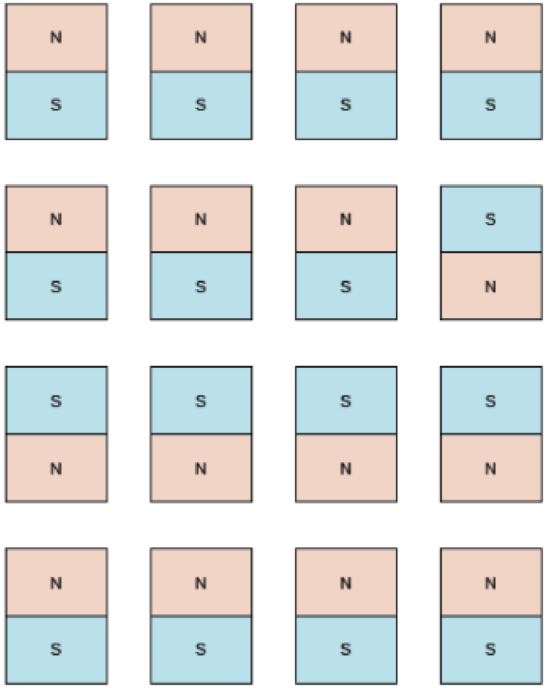
The Ising Model

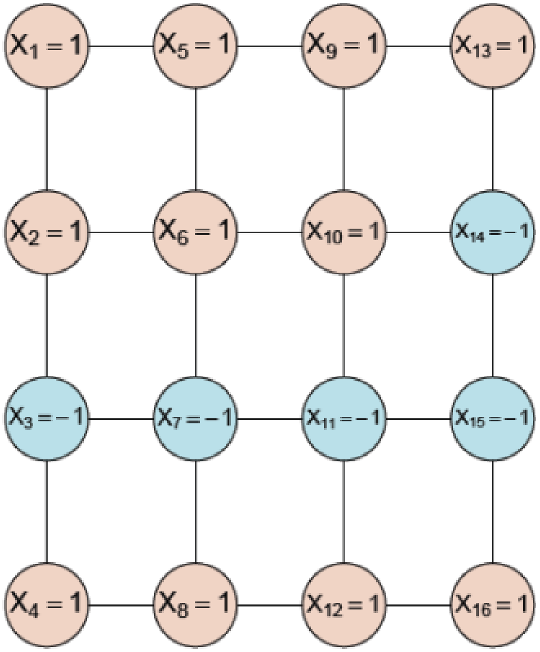


Adopted from the property of Ferromagnetism in physics

Ferromagnetism arises when a collection of atomic spins align such that their associated magnetic moments all point in the same direction, yielding a net magnetic moment which is macroscopic in size.

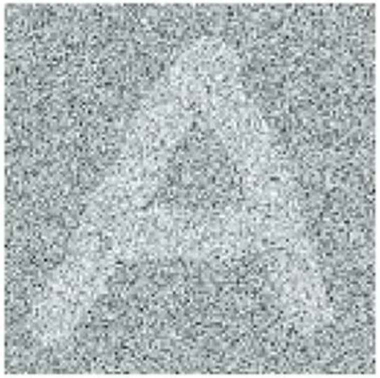
The simplest theoretical description of ferromagnetism is called the Ising model.

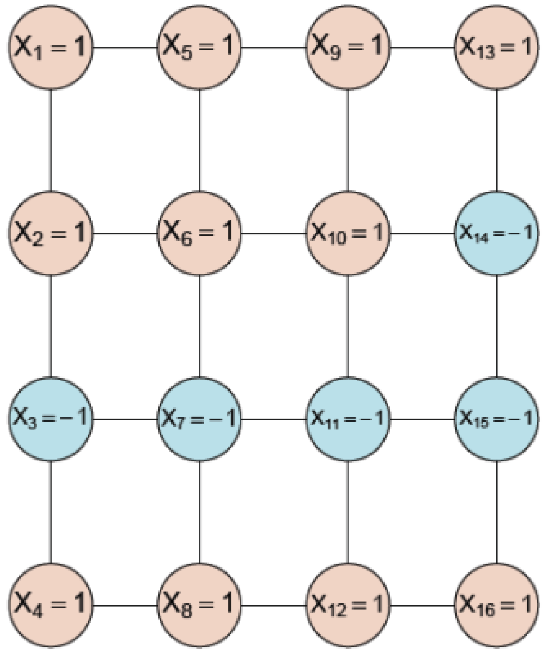




1 variable per unit and connecting neighbouring units

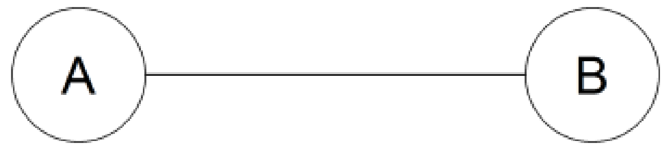
Widely used in image processing





Can you use Bayesian network here?
What will be the direction of causality?

Markov Random Fields as Generating Structure

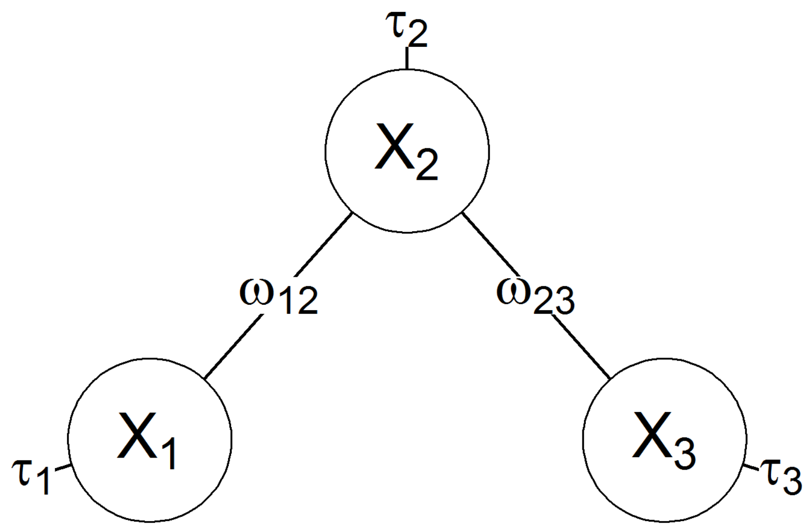


- In the Ising model, we could hold a magnet in some way—$\mathrm{Do}(A)$—which can cause adjacent nodes to "flip" with the same probability if we conditioned on A
  - $\Pr(B \mid \mathrm{Do}(A)) = \Pr(B \mid A)$
  - $\Pr(A \mid \mathrm{Do}(B)) = \Pr(A \mid B)$
- Symmetric relationship that can not be represented in a DAG
- Real relationship that occurs in physics

The Ising Model Probability distribution

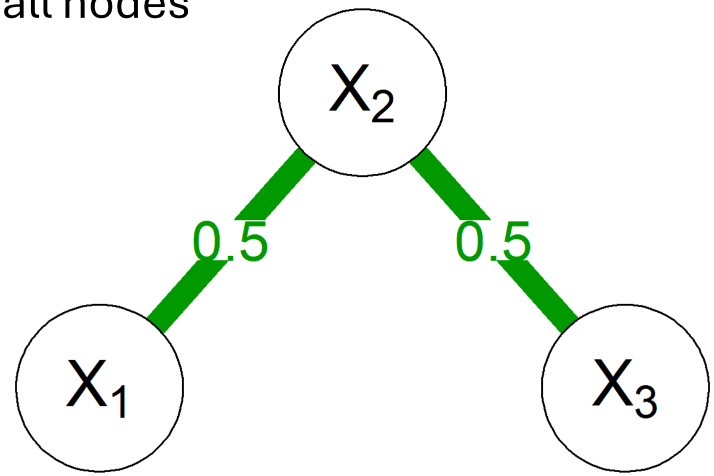
$$\Pr\left(\boldsymbol{X} = \boldsymbol{x}\right) = \frac{1}{Z} \exp\left(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j\right)$$

- All X variables can typically take the values -1 and 1
- τ_i is called the *threshold* parameter and denotes the tendency for node i to be in some state
- ω_{ij} is called the *network* parameter and denotes the preference for nodes i and j to be in the same state
  - Edge weights
- Z is a normalizing constant (partition function) and takes the sum over all possible configurations of $\boldsymbol{X}$:
  - $Z = \sum_{\boldsymbol{x}} \exp\left(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j\right)$



$$\mathbf{\Omega} = \begin{bmatrix} 0 & \omega_{12} & 0 \\ \omega_{12} & 0 & \omega_{23} \\ 0 & \omega_{23} & 0 \end{bmatrix}, \boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

What is the probability that all nodes are 1?



$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0 \end{bmatrix}, \boldsymbol{\tau} = \begin{bmatrix} -0.1 \\ -0.1 \\ -0.1 \end{bmatrix}$$

$$\Pr\left(X = \boldsymbol{x}\right) = \frac{1}{Z}\exp\left(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j\right)$$

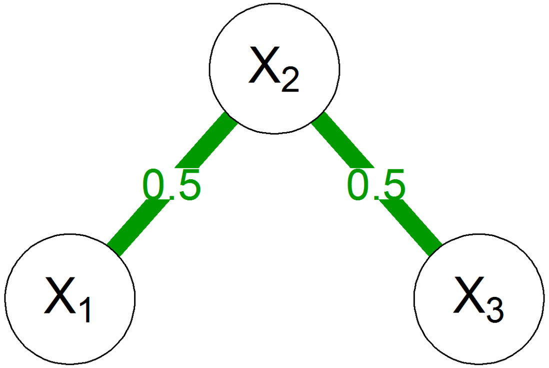
We can compute the unnormalized probability that all nodes are 1:

$$\exp(-0.1 + -0.1 + -0.1 + 0.5 + 0.5) = 2.0138$$

- We can compute the unnormalized probability that all nodes are 1

$$\exp(-0.1 + -0.1 + -0.1 + 0.5 + 0.5) = 2.0138$$

- We will call this the potential for the nodes to be in this state
- Summing the potential of every possible state gives the normalizing constant Z
- Which can then be used to compute the probabilities



x_1	x_2	x_3	Potential	Probability
-1	-1	-1	3.6693	0.3514
1	-1	-1	1.1052	0.1058
-1	1	-1	0.4066	0.0389
1	1	-1	0.9048	0.0866
-1	-1	1	1.1052	0.1058
1	-1	1	0.3329	0.0319
-1	1	1	0.9048	0.0866
1	1	1	2.0138	0.1928
			$Z = 10.4426$	

Gaussian Random Field

Continous Data

If $\boldsymbol{x}$ is not binary but assumed Gaussian we can use a multivariate Gaussian distribution:

$$f(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

- $\boldsymbol{\mu}$ is a vector that encodes the means
- $\boldsymbol{\Sigma}$ is the variance-covariance matrix
- Now we can rearrange:
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})\right)$
  - $f(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right)$

Applications of MRF

Social Network Analysis

MRFs are used to predict missing connections in a social network (e.g., friend recommendations in Facebook or LinkedIn) based on observed relationships

Object Recognition in Computer Vision

Conditional Random Fields (CRFs) are used for semantic segmentation, where each pixel in an image is classified as belonging to an object like "car," "tree," or "road”

Natural Language Processing (NLP) – Part-of-Speech Tagging

Hidden Markov Models (HMMs) (a special case of MRFs) are used in NLP for speech recognition and POS tagging to predict the most likely sequence of word labels

Medical Image Segmentation

MRF-based models help in brain tumor segmentation, where different parts of the brain (gray matter, white matter, cerebrospinal fluid) are identified automatically