Graphical models

Artificial intelligence (CK0031)

Francesco Corona

Graphical models

Graphical models

Graphical models

Belief networks represent independence statements between the variables in a probabilistic model

- BNs are one way to unite probability and graphical representation

Many others exist, all under the wide heading of 'graphical models'

- Each has specific strengths and weaknesses

Whilst not a strict separation, graphical models fall into two classes

- Those useful for modelling
- Those useful for inference

We will survey some of the most popular models from each class

Graphical models (cont.)

Graphical Models (**GMs**) depict independence/dependence relations

- GM classes are particular unions of graph and probability constructs
- The class details the form of independence assumptions represented

Remark

GMs are useful since they provide a framework for studying a wide class of probabilistic models and associated algorithms

- They help to clarify modelling assumptions and provide a unified framework under which inference algorithms can be related

Graphical models (cont.)

All forms of GM have a limited ability to graphically express conditional (in)dependence statements

- BNs are useful for modelling ancestral conditional independence
- Other types are more suited to representing different assumptions

We focus on **Markov networks**, **chain graphs** and **factor graphs**

- There are many more

Graphical models
UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Graphical models (cont.)

We describe the problem environment using a probabilistic model

- Reasoning corresponds to performing probabilistic inference

This is a two-part process:

1. **Modelling**
2. **Inference**

Graphical models
UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Graphical models (cont.)

- **Modelling**: After identifying all potentially relevant variables of a problem environment, we describe how these variables can interact

Remark

Structure assumptions as to the form of the joint probability distribution of all variables (typically, assumptions of independence of variables)

- Each class of graphical model corresponds to a factorisation property of the joint distribution

Graphical models
UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Graphical models (cont.)

- **Inference**: Once the basic assumptions as to how variables interact with each other is formed (i.e. the probabilistic model is built) all questions are answered by performing inference on the distribution

Remark

This can be a computationally non-trivial step so that coupling GMs with accurate inference algorithms is central to graphical modelling

Graphical models (cont.)

Whilst not a strict separation, GMs tend to fall into two broad classes

- Those useful in modelling
- Those useful in representing inference algorithms

For modelling: Belief networks, Markov networks, chain graphs and influence diagrams are some of the most popular

For inference: One 'compiles' a model into a suitable GM for which an algorithm can be readily applied

- Such inference GMs include factor graphs and junction trees

Markov networks

Graphical models

Markov networks

Belief networks correspond to a special kind of factorisation of the joint probability distribution in which each of the factors is itself a distribution

An alternative factorisation is given by

$$p(a,b,c) = \frac{1}{Z}\phi(a,b)\phi(b,c) \quad (1)$$

$\phi(a,b)$ and $\phi(b,c)$ are **potentials** and Z is a constant called **partition function** which ensures normalisation

$$Z = \sum_{a,b,c} \phi(a,b)\phi(b,c) \quad (2)$$

Markov networks (cont.)

Definition

Potentials and joint potentials

A potential is a nonnegative function of variable x, $\phi(x) \geq 0$, and a joint potential is a nonnegative function $\phi(x_1, \dots, x_n)$ of a set of variables

A distribution is a special case of a potential satisfying normalisation

$$\sum_x \phi(x) = 1$$

This holds for continuous variables (summation replaced by integration)

- We use the convention that the ordering of the variables in the potential is not relevant (as for the distribution)
- Joint variables simply index an element of the potential table

Markov networks (cont.)

Definition

Markov network: For a set of variables $\mathcal{X} = \{x_1, \ldots, x_n\}$, a Markov net is defined as a product of potentials on subsets of the variables $\mathcal{X}_c \subseteq \mathcal{X}$

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi_c(\mathcal{X}_c) \quad (3)$$

The constant Z ensures the distribution is normalised

Graphically this is represented by an undirected graph $\mathcal{G}$

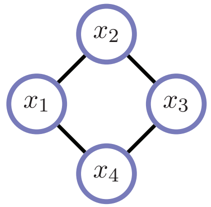
- $\{\mathcal{X}_c\}_{c=1}^{C}$ being the maximal cliques of $\mathcal{G}$

Graphical models

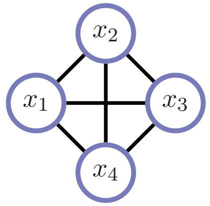
UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)



$\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_1)/Z_a$



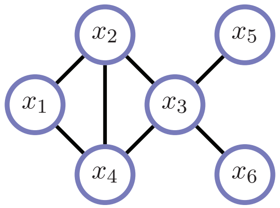
$\phi(x_1, x_2, x_3, x_4)/Z_b$

Graphical models

UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)



$$\phi(x_1, x_2, x_4)\phi(x_2, x_3, x_4)$$
$$\phi(x_3, x_5)\phi(x_3, x_6)/Z_c$$

Graphical models

UFC/DC
AI (CK0031)
2016.2

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)

Definition

Gibbs distribution: A Markov net with strictly positive clique potentials

Definition

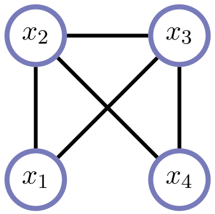
Pairwise Markov network: A Markov net in which the graph contains cliques of size 2 only and potentials defined on each link between vars

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)

MNs are defined as products on maximal cliques of an undirected graph

- Some authors use the term to refer to maximal-cliques also



The maximal cliques are $\{x_1, x_2, x_3\}$ and $\{x_2, x_3, x_4\}$ so that the graph describes a distribution $p(x_1, x_2, x_3, x_4)$

$$p(x_1, x_2, x_3, x_4) = \phi(x_1, x_2, x_3)\phi(x_2, x_3, x_4)/Z$$

In a pairwise MN though potentials are assumed to be over two-cliques, giving $p(x_1, x_2, x_3, x_4) = \frac{1}{Z}\phi(x_1, x_2)\phi(x_1, x_3)\phi(x_2, x_3)\phi(x_2, x_4)\phi(x_3, x_4)$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)

Example

The Boltzmann machine (distribution)

A **Boltzmann machine** is a MN on binary variables, $\text{dom}(x_i) = \{0, 1\}$

$$p(\mathbf{x}) = \frac{1}{Z(\mathbf{w}, b)} \exp\Big(\underbrace{\sum_{i<j} w_{ij} x_i x_j + \sum_i b_i x_i}_{\text{Hamiltonian}}\Big) \quad (4)$$

- The graphical model is an undirected graph with a link between nodes i and j for $w_{ij} \neq 0$

Edge interactions are weights w_{ij} and node potentials are biases b_i

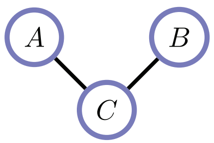
This model has been studied as a basic model of distributed memory

For all but specially constrained $\mathbf{W}$, the graph is multiply connected

- Inference is typically intractable

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov networks (cont.)



$$P(A, B, C) = \frac{1}{Z}\phi_{AC}(A, C)\phi_{BC}(B, C) \quad (5)$$

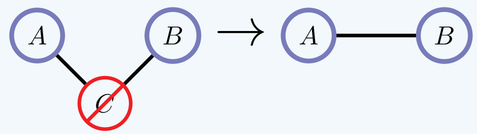
$$\text{with } \frac{1}{Z} = \sum_{A,B,C} \phi_{AC}(A, C)\phi_{BC}(B, C)$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

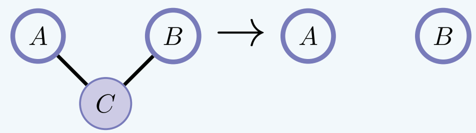
Markov networks (cont.)

Definition

Properties of Markov networks



Marginalising over C makes A and B (graphically) dependent
In general $p(A, B) \neq p(A)p(B)$



Conditioning on C makes A and B independent $A \perp\!\!\!\perp B | C$
$p(A, B|C) = p(A|C)p(B|C)$

Markov properties

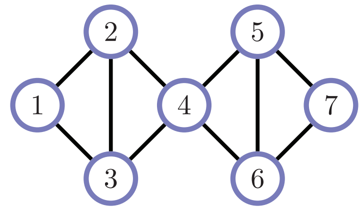
Markov networks

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov properties

We consider somehow informally the properties of Markov networks

We use this graph to show conditional independence properties



Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov properties (cont.)

Let $\phi(1,2,3) \equiv \phi(x_1,x_2,x_3)$, $p(1) \equiv p(x_1)$, $p(2,3) \equiv p(x_2,x_3)$, $\ldots$, etc.

- We divide by potentials and to ensure it is well defined we assume them positive
- For positive potentials, the next local, pairwise and global Markov properties are all equivalent

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov properties (cont.)

Definition

Separation

A subset $\mathcal{S}$ separates a subset $\mathcal{A}$ from a subset $\mathcal{B}$, for disjoint $\mathcal{A}$ and $\mathcal{B}$, if every path from any member of $\mathcal{A}$ to any member of $\mathcal{B}$ passes thru $\mathcal{S}$

- If there are no paths from a member of $\mathcal{A}$ to a member of $\mathcal{B}$ then $\mathcal{A}$ is separated from $\mathcal{B}$

If $\mathcal{S} = \emptyset$, provided no path exists from $\mathcal{A}$ to $\mathcal{B}$, $\mathcal{A}$ and $\mathcal{B}$ are separated

Markov properties (cont.)

Definition 2.1

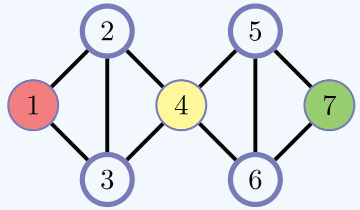
Global Markov property

For disjoint sets of variables $(\mathcal{A}, \mathcal{B}, \mathcal{S})$ where $\mathcal{S}$ separates $\mathcal{A}$ from $\mathcal{B}$ in $\mathcal{G}$, then $\mathcal{A} \perp\!\!\!\perp \mathcal{B} | \mathcal{S}$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

As an example, of the global Markov property consider the following

Example



Are 1 and 7 independent, given 4? Is $1 \perp\!\!\!\perp 7|4$?

$$p(1,7|4) \propto \sum_{2,3,5,6} p(1,2,3,4,5,6,7)$$
$$= \sum_{2,3,5,6} \phi(1,2,3)\phi(2,3,4)\phi(4,5,6)\phi(5,6,7)$$
$$= \Big\{ \sum_{2,3} \phi(1,2,3)\phi(2,3,4) \Big\} \Big\{ \sum_{5,6} \phi(4,5,6)\phi(5,6,7) \Big\}$$
$$\Rightarrow p(1|4)p(7|4)$$

This can be inferred as all paths from node 1 to 7 pass necessarily thru 4

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov properties (cont.)

Pseudocode

An algorithm for independence

The separation property implies an algorithm for deciding $\mathcal{A} \perp\!\!\!\perp \mathcal{B} | \mathcal{S}$

- We simply remove all links that neighbour the set of variables $\mathcal{S}$
- If there is no path from any member of $\mathcal{A}$ to any member of $\mathcal{B}$, then $\mathcal{A} \perp\!\!\!\perp \mathcal{B} | \mathcal{S}$ is true

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov properties (cont.)

For positive potentials, the so-called **local Markov property** holds:

$$p(x|\mathcal{X} \backslash x) = p(x|\text{ne}(x)) \quad (6)$$

When conditioned on its neighbours, x is independent of others

The **pairwise Markov property** holds for non-adjacent vertices x and y

$$x \perp\!\!\!\perp y | \mathcal{X} \backslash \{x, y\} \quad (7)$$

Markov random fields

Markov networks

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Markov random fields

A **Markov random field** (**MRF**) is a set of conditional distributions

- one for each 'indexed' location

Definition

Markov random field

A MRF is defined by a set of distributions $p(x_i|\text{ne}(x_i))$, $i \in \{1, \dots, n\}$ indexes the distributions and $\text{ne}(x_i)$ are the neighbours of variable x_i

- Namely, $\text{ne}(x_i)$ is the subset of variables $x_1, \dots, x_n$ that the distribution of variable x_i depends on
- The term Markov indicates that this is a proper subset of variables

A distribution is a MRF with respect to an undirected graph $\mathcal{G}$ if

$$p(x_i|x_{\setminus i}) = p(x_i|\text{ne}(x_i)) \tag{8}$$

$\text{ne}(x_i)$ are neighbours of x_i according to the undirected graph $\mathcal{G}$

- Notation $\setminus i$ is shorthand for the set of all variables $\mathcal{X}$ excluding variable x_i ($\mathcal{X} \setminus x_i$, in set notation)

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem

Markov networks

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem

An undirected graph $\mathcal{G}$ specifies a set of independence statements

- How to find the most general functional form of the distribution that satisfies the independence statements

Example

A trivial example is graph $x_1 - x_2 - x_3$ from which $x_1 \perp\!\!\!\perp x_3 | x_2$

- From this we must have $p(x_1|x_2, x_3) = p(x_1|x_2)$

$$\begin{aligned} p(x_1, x_2, x_3) &= p(x_1|x_2, x_3)p(x_2, x_3) = p(x_1|x_2)p(x_2, x_3) \\ &= \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3) \end{aligned} \tag{9}$$

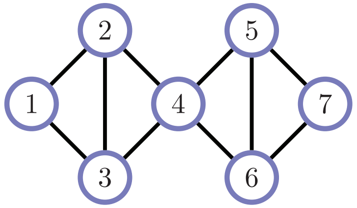
More generally, for any decomposable graph $\mathcal{G}$[1], we can start at the edge and work inwards to reveal that the functional form must be a product of potentials on the cliques of $\mathcal{G}$

[1] Triangulated (Decomposable) Graph: An undirected graph is triangulated if every loop of length 4 or more has a chord. An equivalent term is that the graph is chordal.

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

Start with x_1 and its local Markov statement $x_1 \perp\!\!\!\perp x_4, x_5, x_6, x_7 | x_2, x_3$



$$p(x_1, \dots, x_7) = p(x_1 | x_2, x_3, \cancel{x_4}, \cancel{x_5}, \cancel{x_6}, \cancel{x_7}) p(x_2, x_3, x_4, x_5, x_6, x_7) \qquad (10)$$

Consider x_1 eliminated and move to the neighbours of x_1, x_2 and x_3

From graph, x_1, x_2 and x_3 are independent of x_5, x_6 and x_7 given x_4

$$p(x_1, x_2, x_3 | x_4, x_5, x_6, x_7) = p(x_1, x_2, x_3 | x_4) \qquad (11)$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

$$p(x_1, x_2, x_3 | x_4, x_5, x_6, x_7) = p(x_1, x_2, x_3 | x_4)$$

By summing both sides over x_1, $p(x_2, x_3 | x_4, x_5, x_6, x_7) = p(x_2, x_3 | x_4)$ thus

$$\begin{aligned} p(x_2, x_3, x_4, x_5, x_6, x_7) &= p(x_2, x_3 | x_4, x_5, x_6, x_7) p(x_4, x_5, x_6, x_7) \\ &= p(x_2, x_3 | x_4) p(x_4, x_5, x_6, x_7) \end{aligned}$$

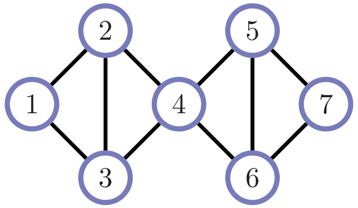
and

$$p(x_1, \dots, x_7) = p(x_1 | x_2, x_3) p(x_2, x_3 | x_4) p(x_4, x_5, x_6, x_7)$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

We eliminated x_2 and x_3 and we move to their neighbour(s), namely x_4



$$p(x_1, \dots, x_7) = p(x_1 | x_2, x_3) p(x_2, x_3 | x_4) p(x_4 | x_5, x_6) p(x_5, x_6 | x_7) p(x_7)$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

The pattern shows that Markov conditions mean that the distribution is expressible as a product of potentials defined on the cliques of the graph

- $\mathcal{G} \Longrightarrow F$ where F is a factorisation into clique potentials on $\mathcal{G}$

The converse is easily shown: That is, given a factorisation F into clique potentials, the Markov conditions on $\mathcal{G}$ are implied

Hence $\mathcal{G} \Longleftrightarrow F$ and it is clear that for any decomposable $\mathcal{G}$, this always holds since we can always work inwards from the edges of the graph

Hammersley-Clifford theorem (cont.)

The Hammersley-Clifford theorem is a stronger result and it shows that this factorisation property holds for any undirected graph, provided that the potentials are positive

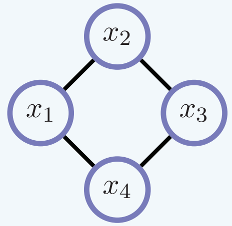
- An informal argument can be made by considering an example

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

Example

Consider the four-cycle $x_1 - x_2 - x_3 - x_4 - x_1$



The theorem states that for positive potentials ϕ, the Markov conditions implied by the graph mean that the distribution must be of the form

$$p(x_1, x_2, x_3, x_4) = \phi_{12}(x_1 x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{41}(x_4, x_1) \qquad (12)$$

It can be shown that for any distribution of this form $x_1 \perp\!\!\!\perp x_3 | x_2, x_4$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

Consider including an additional term that links x_1 to a variable not a member of the cliques that x_1 inhabits

- That is we include a term $\phi_{13}(x_1, x_3)$

Our aim is to show that a distribution of the form

$$p(x_1, x_2, x_3, x_4) = \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{41}(x_4, x_1)\phi_{13}(x_1, x_3) \qquad (13)$$

cannot satisfy the Markov property $x_1 \perp\!\!\!\perp x_3 | x_2, x_4$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

$$p(x_1, x_2, x_3, x_4) = \frac{\phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{41}(x_4, x_1)\phi_{13}(x_1, x_3)}{\sum_{x_1} \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{41}(x_4, x_1)\phi_{13}(x_1, x_3)} = \frac{\phi_{12}(x_1, x_2)\phi_{41}(x_4, x_1)\phi_{13}(x_1, x_3)}{\sum_{x_1} \phi_{12}(x_1, x_2)\phi_{41}(x_4, x_1)\phi_{13}(x_1, x_3)} \qquad (14)$$

If we assume that potential $\phi_{13}(x_1, x_3)$ is weakly dependent on x_1 and x_3,

$$\phi_{13}(x_1, x_3) = 1 + \varepsilon\psi(x_1, x_3), \quad \text{with } \varepsilon << 1 \qquad (15)$$

Hammersley-Clifford theorem (cont.)

$$p(x_1|x_2,x_3,x_4) = \frac{\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}(1+\varepsilon\psi(x_1,x_3)) \Big(1+\varepsilon \underbrace{\frac{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)\psi(x_1,x_3)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}}_{f}\Big)^{-1} \quad (16)$$

By expanding $(1+\varepsilon f)^{-1} = 1 - \varepsilon f + \mathcal{O}(\varepsilon^2)$ and retaining only terms that are first order in ε, we obtain

$$p(x_1|x_2,x_3,x_4) = \frac{\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)} \Big(1+\varepsilon\Big[\psi(x_1,x_3) - \frac{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)\psi(x_1,x_3)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}\Big]\Big)^{-1} + \mathcal{O}(\varepsilon^2) \quad (17)$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

$$p(x_1|x_2,x_3,x_4) = \frac{\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)} \Big(1+\varepsilon\Big[\psi(x_1,x_3) - \frac{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)\psi(x_1,x_3)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)}\Big]\Big) + \mathcal{O}(\varepsilon^2) \quad (18)$$

- The first factor is independent of x_3 as required by the Markov condition, for $\varepsilon \neq 0$ the second term varies as a function of x_3

The reason is that one can always find a function $\psi(x_1,x_3)$ for which

$$\psi(x_1,x_3) \neq \frac{\sum_{x_1}\psi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)\psi(x_1,x_3)}{\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{41}(x_4,x_1)} \quad (19)$$

since the term $\psi(x_1,x_3)$ on the left is functionally dependent on x_1 whereas the term on the right is not a function of x_1

- Hence, the only way to ensure that the Markov condition holds is if $\varepsilon = 0$ for which there is no connection between x_1 and x_3

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

The Hammersley-Clifford theorem also helps resolve other questions

- When a set of positive local conditional distributions $p(x_i|\text{pa}(x_i))$ does ever form a consistent joint distribution $p(x_1,\dots,x_n)$?

Each local conditional distribution $p(x_i|\text{pa}(x_i))$ corresponds to a factor on the set of variables $\{x_i|\text{pa}(x_i)\}$, so we must include it in the joint

The MN can form a joint distribution consistent with the local conditional distributions iff $p(x_1,\dots,x_n)$ factorises according to

$$p(x_1,\dots,x_n) = \frac{1}{Z}\exp\Big(-\sum_c V_c(\mathcal{X}_c)\Big) \quad (20)$$

The sum is over all cliques and $V_c(\mathcal{X}_c)$ is a real function defined over all the variables in the clique indexed by c

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

$$p(x_1,\dots,x_n) = \frac{1}{Z}\exp\Big(-\sum_c V_c(\mathcal{X}_c)\Big)$$

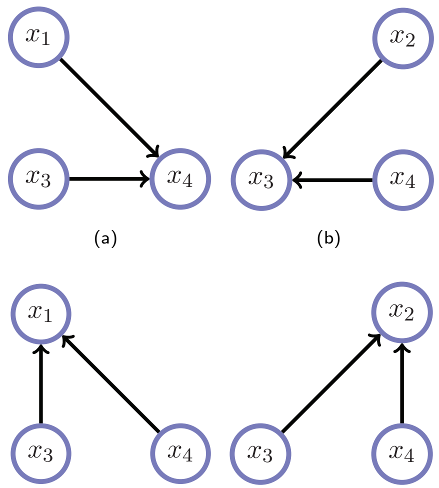
The equation is equivalent to $\prod_c \phi(\mathcal{X}_c)$, namely a Markov network

- On positive cliques potentials

The graph over which the cliques are defined is an undirected graph

- This graph is constructed by taking each local conditional distribution $p(x_i|\text{pa}(x_i))$ and drawing a clique on $\{x_i,\text{pa}(x_i)\}$

This is then repeated over all the local conditional distributions



(a) (b) (c) (d)

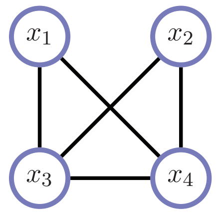
Local conditional distributions: No distribution is implied for the parents

- In (a) we are given the conditional $p(x_4|x_1, x_3)$: One should not read from graph that we imply x_1 and x_3 are marginally independent

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

The Markov network consistent with the local distributions



(e)

If the local distributions are positive, b Hammersley-Clifford theorem

- then the only joint distribution that can be consistent with the local distributions must be Gibbs with structure given by (e)

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Hammersley-Clifford theorem (cont.)

Remark

The HC theorem does not mean that, given a set of conditional distributions, we can always form a consistent joint distribution from them, rather it states what the functional form of a joint distribution has to be for the conditionals to be consistent with it

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Conditional independence using Markov networks

Markov networks

Conditional independence using Markov networks

For $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ each being a collection of variables, we discussed an algorithm to determine if $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$ in the case of belief networks

Remark

'*For every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, check every path U between x and y, a path U is said to be* **blocked** *if there is a node w on U such that either*:

- *w is a collider and neither w nor any of its descendants is in $\mathcal{Z}$*
- *w is not a collider on U and w is in $\mathcal{Z}$*

If all such paths are blocked, then $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$

If the sets $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$, then they are independent conditional on $\mathcal{Z}$ in all distributions such a graph can represent

We can now highlight an alternative and more general method

- Both directed and undirected graphs

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Conditional independence using MNs (cont.)

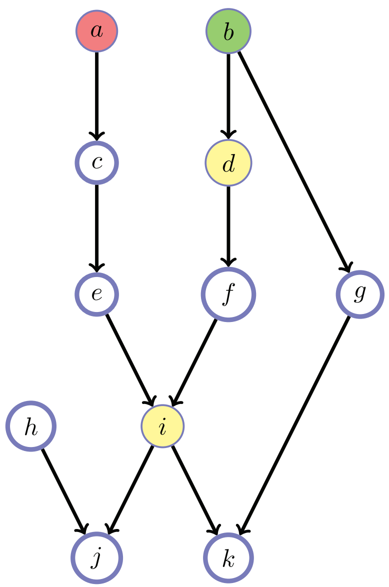
Pseudocode

Ascertaining independence in Markov and belief networks

For MNs only the final separation criterion needs to be applied

- **Ancestral graph**: Identify the ancestors $\mathcal{A}$ of nodes $\mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}$ but remove all other nodes which are not in $\mathcal{A}$ together with any edge in or out of such nodes
- **Moralisation**: Add a link between any two remaining nodes which have a common child, but are not already connected by an arrow, then remove remaining arrowheads
- **Separation**: Remove links neighbouring $\mathcal{Z}$ and in the undirected graph so constructed, look for a path which joins a node in $\mathcal{X}$ to one in $\mathcal{Y}$, then if there is no such path deduce that $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$

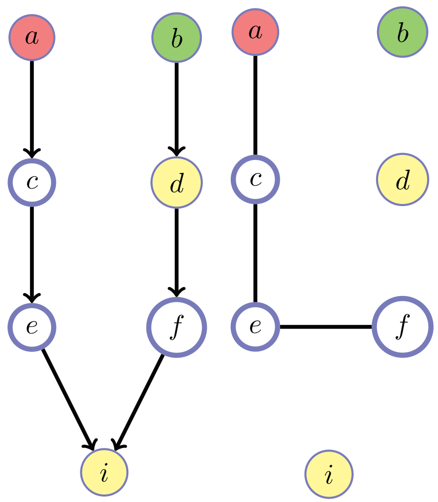
Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models



(a)

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Conditional independence using MNs (cont.)



(b) (c)

Conditional independence using MNs (cont.)

The ancestral step in the procedure for belief networks is intuitive

- Given a set of nodes $\mathcal{X}$ and their ancestors $\mathcal{A}$, the remaining nodes $\mathcal{D}$ for a contribution to the distribution of form $p(\mathcal{D}|\mathcal{X},\mathcal{A})p(\mathcal{X},\mathcal{A})$
- Summing over $\mathcal{D}$ has the effect of removing these vars from DAG

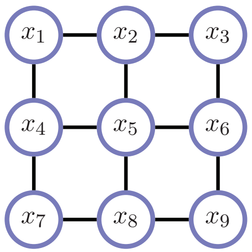
Lattice models
Markov networks

Lattice models

Undirected models have a history in different branches of science

- Especially statistical mechanics on lattices and models in visual processing that encourage neighbours to be in the same states

Consider the model in which our desire is that states of binary variables $\{x_i\}_{i=1}^{9}$ on a lattice should prefer neighbours to be in the same state



$$p(x_1,\dots,x_9) = \frac{1}{Z}\prod_{i\sim j}\phi_{ij}(x_i,x_j) \quad (21)$$

$i \sim j$ denotes sets of indices where j are neighbours of i in the undirected graph

Ising models

A set of potentials that encourages neighbours to have the same state is

$$\phi_{ij}(x_i,x_j) = \exp\Big(-\frac{1}{2T}(x_i-x_j)^2\Big), \quad \text{with } x_i \in \{-1,+1\} \qquad (22)$$

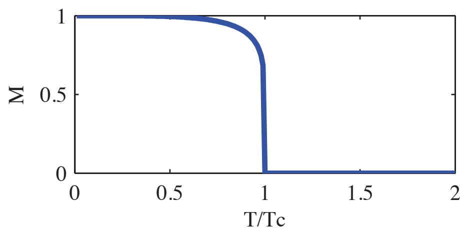
This corresponds to a well-known model for the physics of magnetic systems, the **Ising model**, which consists of 'mini-magnets' which prefer to be aligned in the same state, depending on the temperature T

- High T: Variables behave independently, so that no global magnetisation appears
- Low T: Preference for neighbours to become aligned, generating a strong macro-magnet

Ising models (cont.)

Remarkably, one can show a behaviour in a large 2-dimensional lattices

- Below the so-called Curie-temperature $T_C \simeq 2.27$ for ± 1 variables, the systems admits a phase change in that a large fraction of the variables become aligned
- above T_C the variables remain unaligned, on average



Average alignment of variables

$$M = \frac{1}{N}\Big|\sum_{i=1}^{N} x_i\Big|$$

Onsager magnetisation

As T decreases towards the critical value T_C, a phase transition occurs in which a large fraction of the variables become aligned in the same state

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Ising models (cont.)

Global coherence effects such as this that arise from weak local constraints are present in systems that admit emergent behaviour

Similar local constraints are common in image denoising algos, under the assumption that noise has no local spatial coherence, whilst 'signal' does

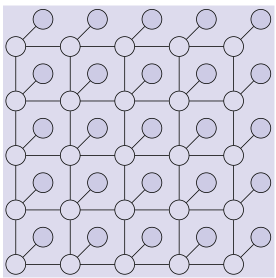
Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Ising models (cont.)

Example

Cleaning up images: Consider a binary image on pixels $x_i \in \{-1, +1\}$ with $i = 1, \dots, D$ and observe a noisy version y_i of each pixel x_i in which the state of $y_i \in \{-1, +1\}$ is opposite to x_i with some probability

Clean up the observed dirt image $\mathcal{Y}$ and find most likely clean image $\mathcal{X}$



- Filled nodes are observed noisy pixels
- Unshaded nodes are latent clean pixels

$$p(\mathcal{X}, \mathcal{Y}) = \frac{1}{Z}\Big[\prod_{i=1}^{D} \phi(x_i, y_i)\Big]\Big[\prod_{i \sim j} \psi(x_i, x_j)\Big]$$

$$\text{with} \begin{cases} \phi(x_i, x_j) = \exp(\beta x_i x_j) \\ \psi(x_i, x_j) = \exp(\alpha x_i x_j) \end{cases}$$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Ising models (cont.)

$i \sim j$ is the set of unobserved (latent) variables that are neighbours

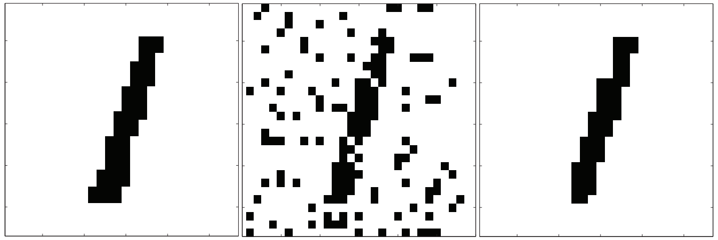
- Potential $\phi(x_i, y_i)$ encourages noisy and clean pixels to be in the same state
- Potential $\psi(x_i, x_j)$ encourages neighbouring pixels to be in the same state

To find the most likely clean image, we need to compute

$$\arg\max_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}) = \arg\max_{\mathcal{X}} p(\mathcal{X}, \mathcal{Y}) \quad (23)$$

It's a difficult task, but can be approximated with iterative methods

Ising models (cont.)



On the left is the clean image from which a noisy corrupted image $\mathcal{Y}$ is formed in the middle and on the right the most likely restored image $\mathcal{X}$

Parameter β can be set from knowledge of corruption probability p_{corrupt}

$$p(y_i \neq x_i | x_i) = \sigma(-2\beta), \quad \text{so } \beta = -\frac{1}{2}\sigma^{-1}(p_{\text{corrupt}})$$

Parameter α is more complex, since relating $p(x_i = x_j)$ to α is not easy

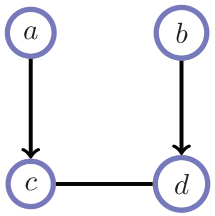
- (here we set $\alpha = 10$)

Chain graphical models

Graphical models

Chain graphical models

Chain graphs (**CG**) contain both directed and undirected links



To develop the intuition consider the graph

The terms we can unambiguously specify are $p(a)$ and $p(b)$, since there is no mixed interaction of directed/undirected edges at a and b nodes

By probability, we must have

$$p(a, b, c, d) = p(a)p(b)p(c, d|a, b) \tag{24}$$

From graph, we expect the interpretation to be

$$p(c, d|a, b) = \phi(c, d)p(c|a)p(d|b) \tag{25}$$

Chain graphical models

To ensure normalisation and to retain generality, we interpret this as

$$p(c, d|a, b) = \phi(c, d)p(c|a)pd|b)\phi(a, b) \tag{26}$$

with $\phi(a, b) \equiv \Big(\sum_{c,d} \phi(c, d)p(c|a)p(d|b)\Big)^{-1}$

We can interpret the CG as a DAG over the chain components

Chain graphical models (cont.)

Definition

Chain component: Chain components of graph $\mathcal{G}$ are obtained by

1. Form a graph $\mathcal{G}'$ with directed edges removed from $\mathcal{G}$
2. Each connected component in $\mathcal{G}'$ constitutes a component

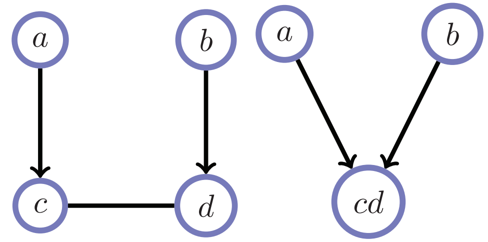
Each chain component represents a distribution over the variables of the component, conditioned on the parental components

The conditional distribution is itself a product over the cliques of the undirected component and moralised parental components, including also a factor to ensure normalisation over the chain component

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Chain graphical models (cont.)

The chain components are identified by deleting the directed edges and identifying the remaining connected components

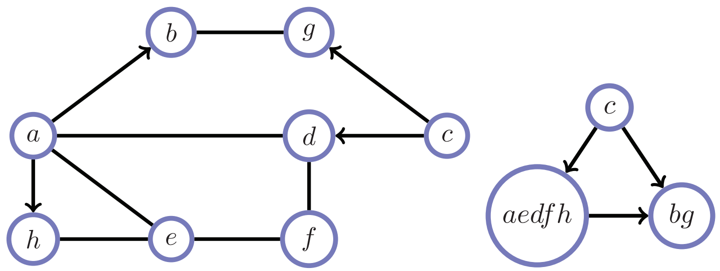


(a) (b)

- Case a) Chain components are (a), (b) and (c, d), which can be written as a BN on the cluster variables in Case b)

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Chain graphical models (cont.)



(c) (d)

- Case c) Chain components are (a, e, d, f, h), (b, g) and (c), which has the cluster BN representation in Case d)

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Chain graphical models (cont.)

Definition

Chain graph distribution

The distribution associated with a chain graph G is found by first identifying the chain components, τ and associated vars $\mathcal{X}_\tau$, then

$$p(x) = \prod_\tau p(\mathcal{X}_\tau | \text{pa}(\mathcal{X}_\tau))$$
$$p(\mathcal{X}_\tau | \text{pa}(\mathcal{X}_\tau)) \propto \prod_{d \in \mathcal{D}_\tau} p(x_d | \text{pa}(x_d)) \prod_{c \in \mathcal{C}_\tau} \phi(\mathcal{X}_c) \tag{27}$$

- $\mathcal{C}_\tau$ denotes the union of the cliques in component τ with ϕ being the associated functions defined on each clique
- $\mathcal{D}_\tau$ is the set of variables in component τ that correspond to directed terms $p(x_d|\text{pa}(x_d))$

The proportionality factor is determined by the usual constraint

- The distribution sums to 1

Chain graphical models (cont.)

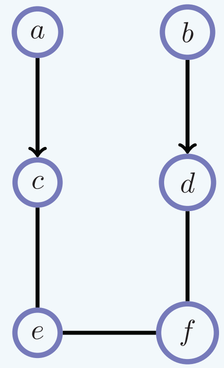
- BNs are CGs in which the connected components are singletons
- MNs are CGs in which the chain components are simply the connected components of the undirected graph

Remark

CGs can be useful as they are more telling of conditional independence statements than either belief networks or Markov networks alone

Chain graphical models (cont.)

Example



Consider the chain graph above with chain component decomposition

$$p(a, b, c, d, e, f) = p(a)p(b)p(c, d, e, f|a, b) \quad (28)$$

Chain graphical models (cont.)

$$p(a, b, c, d, e, f) = p(a)p(b) \underbrace{p(c, d, e, f|a, b)}_{p(c|a)\phi(c,e)\phi(e,f)\phi(d,f)p(d|b)\phi(a,b)}$$

The normalisation requirement is given by the expression

$$\phi(a, b) \equiv \Big(\sum_{c,d,e,f} p(c|a)\phi(c, e)\phi(e, f)\phi(d, f)p(d|b) \Big)^{-1} \quad (29)$$

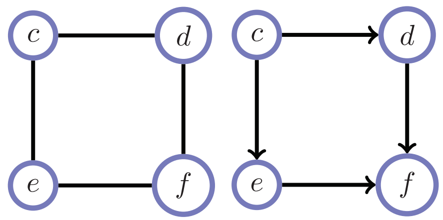
The marginal $p(c, d, e, f)$ is given by the expression

$$\phi(c, e)\phi(e, f)\phi(d, f) \underbrace{\sum_{a,b} \phi(a, b)p(a)p(b)p(c|a)p(d|b)}_{\phi(c,d)} \quad (30)$$

Since the marginal of $p(c, d, e, f)$ is an undirected 4-cycle, no DAG can express the conditional independence statements in $p(c, d, e, f)$

Similarly, no undirected distribution on the same skeleton could express that a and b are independent (unconditionally, $p(a, b) = p(a)p(b)$)

Chain graphical models (cont.)



(a) (b)

Factor graphs
Graphical models

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Factor graphs

Factor graphs (**FGs**) are mainly used as part of inference algorithms

Definition

Factor graphs: Given a function

$$f(x_1, \dots, x_n) = \prod_i \psi_i(\mathcal{X}_i), \quad (31)$$

the factor graph has a node (represented by a square) for each factor ψ_i and a variable node (represented by a circle) for each variable x_j

- For each $x_j \in \mathcal{X}_i$ an undirected link is made between factor ψ_i and variable x_j

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Factor graphs (cont.)

When used to represent a distribution of the following form

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_i \psi_i(\mathcal{X}_i) \quad (32)$$

a normalisation constant $Z = \sum_{\mathcal{X}} \prod_i \psi_i(\mathcal{X}_i)$ is assumed

- $\mathcal{X}$ represents all variables in the distribution

Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

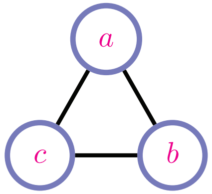
Factor graphs (cont.)

Given a factor $\psi(\mathcal{X}_i)$ which is a conditional distribution $p(x_i|\text{pa}(x_i))$

- We may use a directed links from parents to the factor node and a directed link from the factor node to the child x_i
- This has the same structure as an (undirected) FG but it preserves the information that the factors are distributions

Factor graphs (cont.)

FGs are useful since they preserve more information about the form of the distro than either a Bayes or a Markov network or chain graphs alone



Consider the distribution

$$p(a, b, c) = \phi(a, b)\phi(a, c)\phi(b, c) \quad (33)$$

As a MN, this must have a single clique

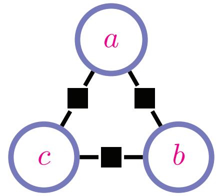
- Though the graph could equally represent some unfactored clique potential $\phi(a, b, c)$

The factorised structure in the clique is lost

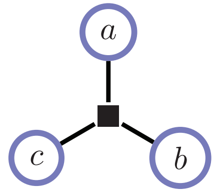
Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Factor graphs (cont.)

A FG more precisely conveys the form of distribution equation $\phi(a, b)\phi(b, c)\phi(c, a)$



An unfactored clique potential $\phi(a, b, c)$ is represented by this other FG depiction



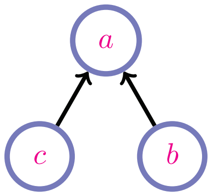
Remark

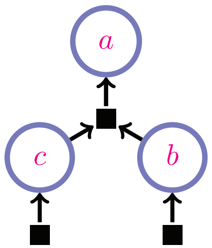
Different FGs can have the same MN since info regarding the structure of the clique potential is lost in the MN

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Factor graphs (cont.)

For a BN, one can represent this using a standard undirected FG, though more information about the independence is preserved by using a directed FG





Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Conditional independence

Factor graphs

Conditional independence in factor graphs

Conditional independence questions can be addressed using a rule which works with directed, undirected and partially directed FGs

To determine whether two variables are independent given a set of conditioned variables, consider all paths connecting the two variables

- If all paths are blocked, the variables are conditionally independent

A path is blocked if one or more of the following conditions is satisfied:

- One of the variables in the path is in the conditioning set
- One of the variables or factors in the path has two incoming edges that are part of the path (variable or factor collider), and neither the variable or factor nor any of its descendants are in the conditioning set

Expressiveness of graphical models

Graphical models

Expressiveness of graphical models

Directed distributions can be represented as undirected distributions

- One can associate each (normalised) factor of the joint distribution with a potential

Example

Distribution $p(a|b)p(b|c)p(c)$ can be factored as $\phi(a,b)\phi(b,c)$, where

- $\phi(a,b) = p(a|b)$
- $\phi(b,c) = p(b|c)p(c)$
- $Z = 1$

Hence every BN can be represented as some MN by a simple identification of the factors in the distributions

Expressiveness of graphical models (cont.)

However, in general, the associated undirected graph (that is, the moralised directed graph) will contain additional links

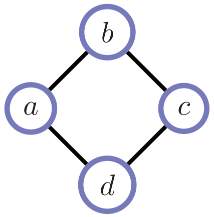
- Independence information can be lost

Example

- The MN of $p(c|a,b)p(a)p(b)$ is a single clique $\phi(a,b,c)$ from which one cannot graphically infer that $a \perp\!\!\!\perp b$

Expressiveness of graphical models (cont.)

The converse question is whether every undirected model can be represented by a BN with a readily derived link structure



In this case, there is no directed model with the same link structure that can express the (in)dependencies in the undirected graph

Naturally, every probability distribution can be represented by some BN

- It may not necessarily have a simple structure
- It may not be a 'fully connected' cascade style graph

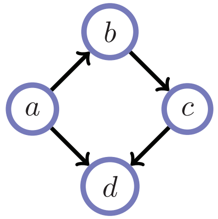
Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Expressiveness of graphical models (cont.)

In this sense the DAG cannot always graphically represent the independence properties that hold for the undirected distribution

Every DAG with the same structure as the undirected model must have a situation where two arrows will point to a node, such as node d

- (otherwise one would have a cyclic graph)



Summing over the states of variable d will leave a DAG on the variables a, b, c with

- no link between a and c

This cannot represent the undirected model since when one marginalises over d this adds a link between a and c

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Expressiveness of graphical models (cont.)

Definition

Independence maps

A graph is an **independence map** (**I-map**) of a given distribution P if every conditional independence statement that one can derive from the graph $\mathcal{G}$ is true in the distribution P

$$\mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_G \implies \mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_P \quad (34)$$

for all disjoint sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$

A graph is a **dependence map** (**D-map**) of a given distribution P if every conditional independence statement that one can derive from P is true on $\mathcal{G}$

$$\mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_G \impliedby \mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_p \quad (35)$$

for all disjoint sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$

Graphical models
UFC/DC
AI (CK0031)
2016.2
Graphical models
Markov networks
Markov properties
Markov random fields
Hammersley-Clifford theorem
Conditional independence using Markov networks
Lattice models
Chain graphical models
Factor graphs
Conditional independence
Expressiveness of graphical models

Expressiveness of graphical models (cont.)

Definition

A graph $\mathcal{G}$ which is both an I-map and a D-map is called a **perfect map**

$$\mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_G \iff \mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}_P \quad (36)$$

for all disjoint sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$

- The set of all conditional independence and dependence statements expressible in the graph $\mathcal{G}$ are consistent with P, and vice versa