Probabilistic Graphical Models

Undirected Models

Siamak Ravanbakhsh

Fall 2019

Learning objectives

Markov networks:

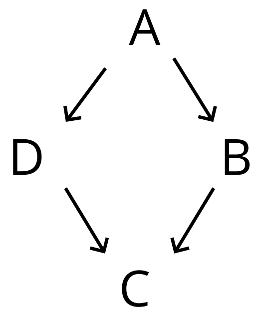
- how it represents a prob. dist.
- independence assumptions
- factorization
- representations:
  - factor-graph
  - log-linear models

Hammersley-Clifford theorem

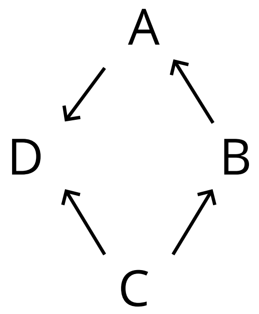
Challenge

Given the following set of CIs draw their DAG

$$\mathcal{I}(P) = \{(A \perp C \mid B, D), (D \perp B \mid A, C)\}$$



OR

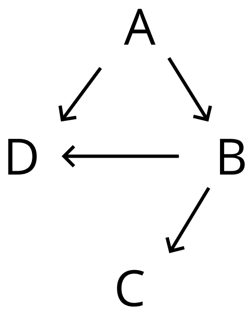


?

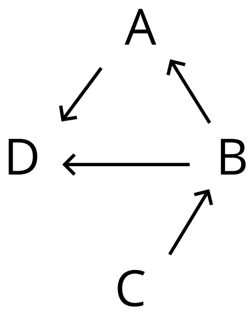
Challenge

Given the following set of CIs draw their DAG

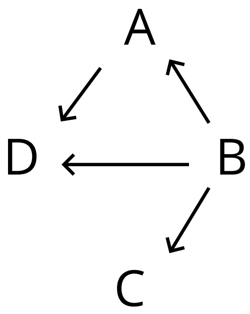
$$\mathcal{I}(P) = \{(A \perp C \mid B, D), (D \perp B \mid A, C)\}$$



OR



OR



?

Challenge

Given the following set of CIs draw their DAG

$$\mathcal{I}(P) = \{(A \perp C \mid B, D), (D \perp B \mid A, C)\}$$

a DAG cannot be a P-map for P

an undirected model can!

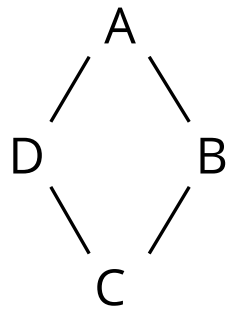
Challenge

Given the following set of CIs draw their DAG

$$\mathcal{I}(P) = \{(A \perp C \mid B, D), (D \perp B \mid A, C)\}$$

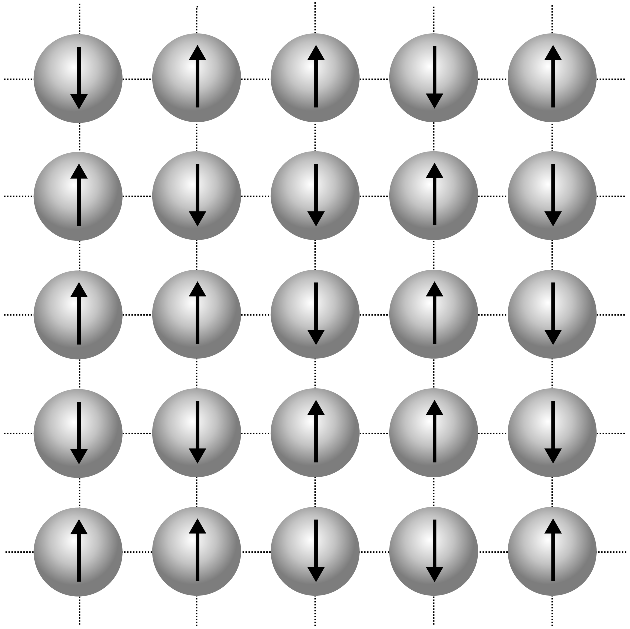
a DAG cannot be a P-map for P

an undirected model can!



Motivation

Statistical physics: **Ising model** of ferromagnetism

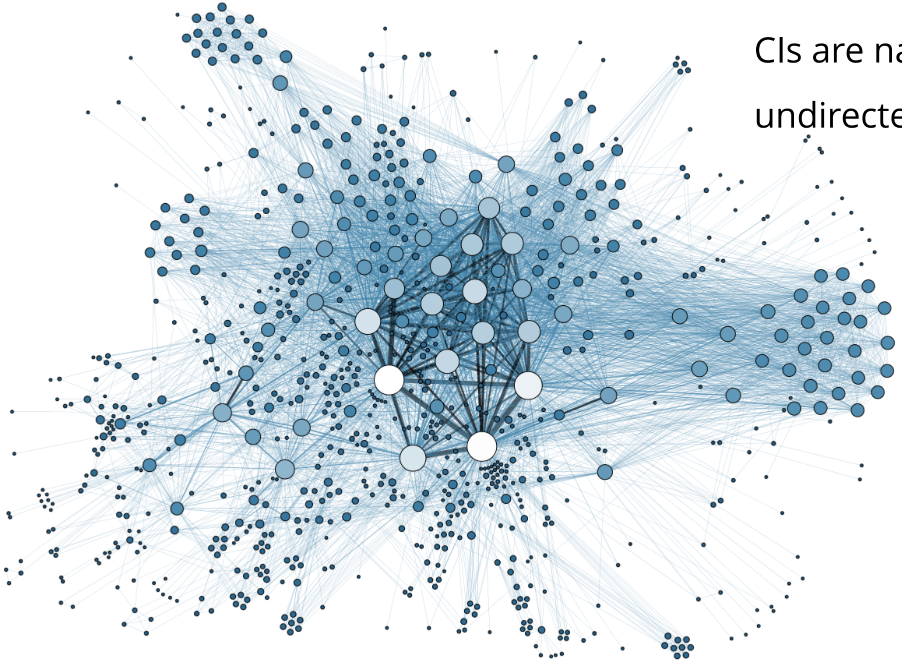


CIs are naturally expressed using an undirected model

Image: https://web.stanford.edu/~peastman/statmech/phasetransitions.html

Motivation

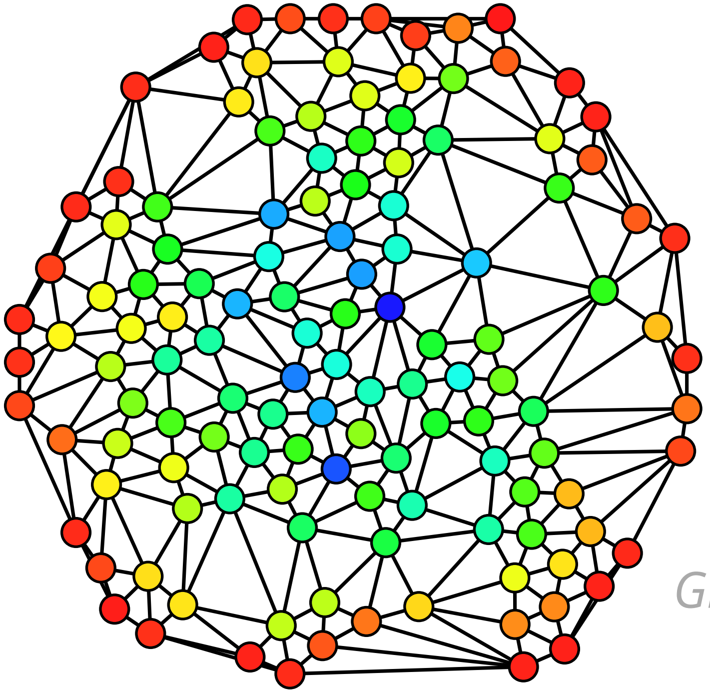
Social sciences



CIs are naturally expressed using an undirected model

Motivation

Combinatorial problems



CIs are naturally expressed using an undirected model

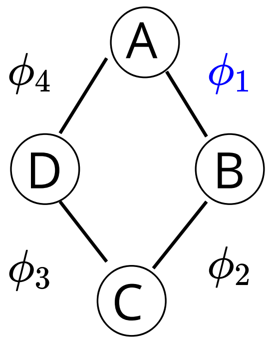
Graph coloring

Factorization in Markov networks

$$P(A,B,C,D) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(A,D)$$

$Z = \sum_{a,b,c,d} \phi_1(a,b)\phi_2(b,c)\phi_3(c,d)\phi_4(a,d)$

is a normalization constant (*partition function*)



$\phi_1 : Val(A,B) \to [0,+\infty)$ is called a factor (*potential*)

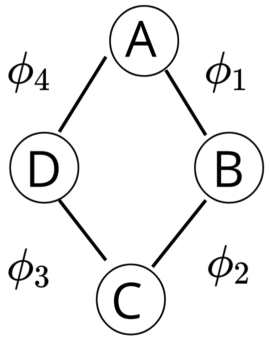
MRF: Conditional Independencies

$$P(A,B,C,D) = \left(\frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\right)\phi_3(C,D)\phi_4(A,D)$$

$f(B,A,C)$ $\quad$ $g(D,A,C)$

assignment (?)

$$P \models (B \perp D \mid A, C)$$



$$P(A,B,C,D) = \left(\frac{1}{Z}\phi_1(A,B)\phi_2(A,D)\right)\phi_3(C,D)\phi_4(B,C)$$

$$P \models (A \perp C \mid B, D)$$

Product of factors

$$P(A,B,C,D) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\phi_3(C,D)\phi_4(A,D)$$

$\phi_1 : Val(A,B) \rightarrow \Re^+$

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

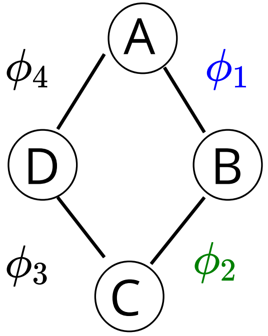
$\phi_2 : Val(B,C) \rightarrow \Re^+$

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

$\psi(A,B,C) : Val(A,B,C) \rightarrow \Re^+$

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

$Val(A) \times Val(B) \times Val(C)$ *similar to a 3D tensor*



Q: Do factors represent marginals?

Simplified example: $P(A,B,C) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)$

ϕ_1

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

ϕ_2

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

$P(A,B,C) \times Z$

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

$Z = .25 + .35 + \ldots = 1.55$

Marginal probabilities:

$P(a^1, b^1) = (.25 + .35)/Z \approx .38$

$P(a^1, b^2) = (.08 + .16)/Z \approx .15$

Compare to ϕ_1

$\phi_1(a^1, b^1) = .5$

$\phi_1(a^1, b^2) = .8$

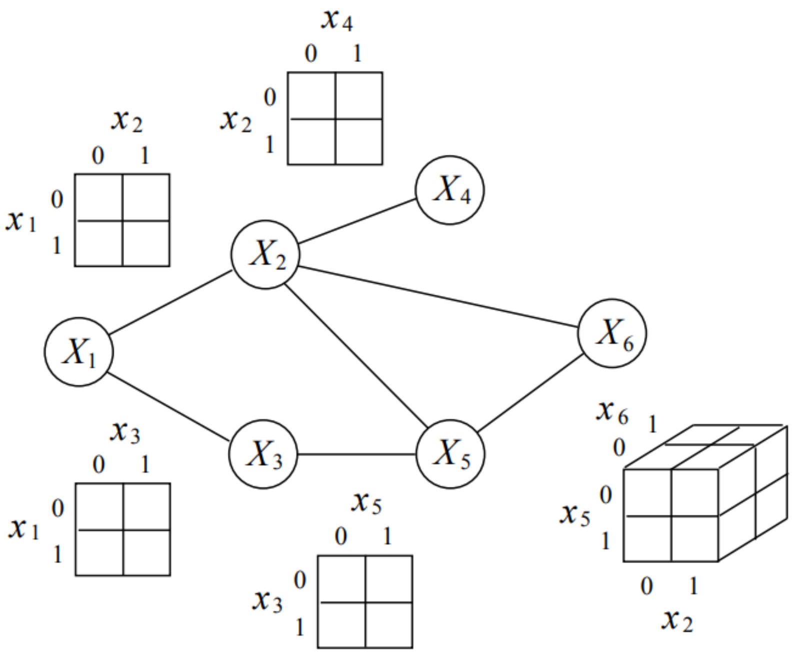
Factorization: general form

P factorizes *over the cliques*

$$P(\mathbf{X}) = \frac{1}{Z} \prod_k \phi_k(\mathbf{D}_k)$$

Gibbs distribution

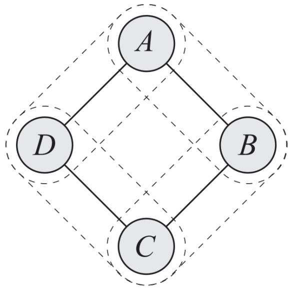
Can always convert to factorization over *maximal cliques*



Factorization: general form

P factorizes *over cliques*

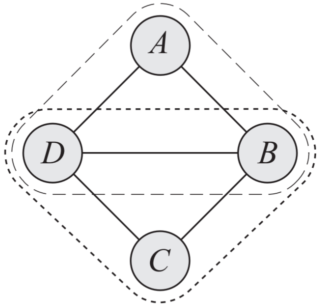
$P(\mathbf{X}) = \frac{1}{Z} \prod_k \phi_k(\mathbf{D}_k)$



Rewrite as factorization over maximal *cliques*

- original form of P

$$P(A,B,C,D) = \phi_1(A,B)\phi_2(A,D)\phi_3(B,D)\phi_4(C,D)\phi_5(B,C)$$



- factorized over cliques

$$P(A,B,C,D) = \psi_1(A,B,C)\psi_2(B,C,D)$$

Factorized form: directed vs undirected

Markov Networks:

$$P(\mathbf{X}) = \frac{1}{Z} \prod_k \phi_k(\mathbf{D}_k)$$

Bayesian Networks:

$$P(\mathbf{X}) = \prod_k P(X_i \mid Pa_{X_i})$$

- No *partition function*
- Each factor is a *cond. distribution*
- One factor per variable

Conditioning on the evidence

given $P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{D}_k)$, how to obtain $P(\mathbf{X} \mid U = u)$?

fix the evidence in the relevant factors $P(\mathbf{X} \mid U = u) \propto \prod_k \phi_k[U = u]$

reduced factor

$\phi_k(A, B, C)$

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

`conditioned on` $C = c^1$ →

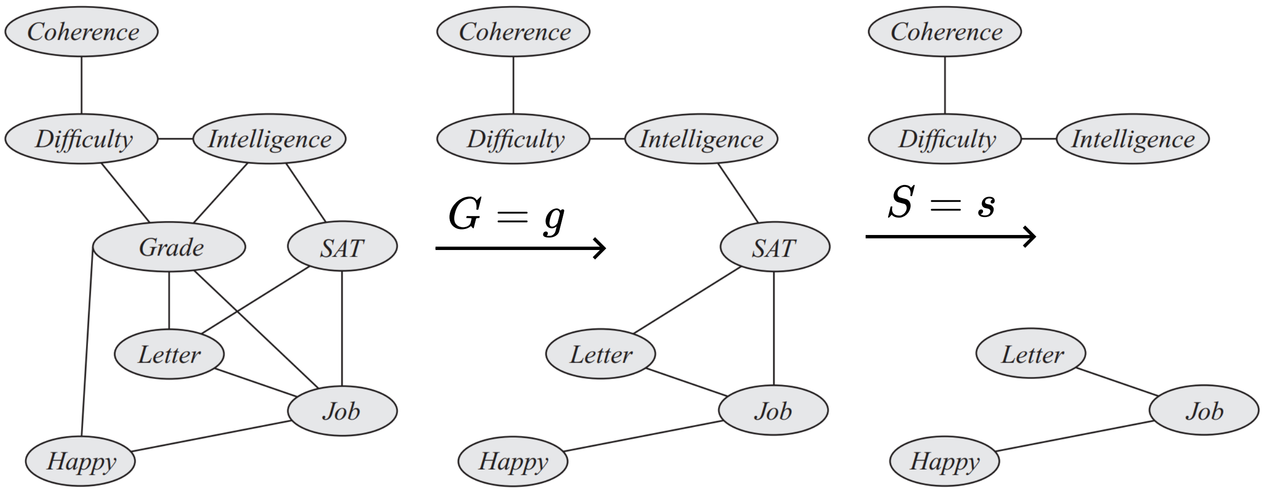
$\phi_k[C = c]$

a^1	b^1	c^1	0.25
a^1	b^2	c^1	0.08
a^2	b^1	c^1	0.05
a^2	b^2	c^1	0
a^3	b^1	c^1	0.15
a^3	b^2	c^1	0.09

Conditioning on the evidence

effect on the graphical model

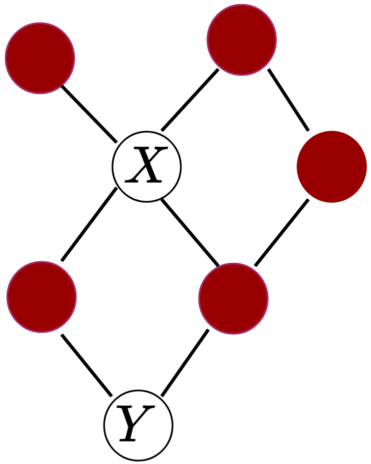
- cannot create new dependencies
- compare this to colliders in **Bayes-nets**



Pairwise conditional independencies

Non-adjacent nodes are independent given everything else

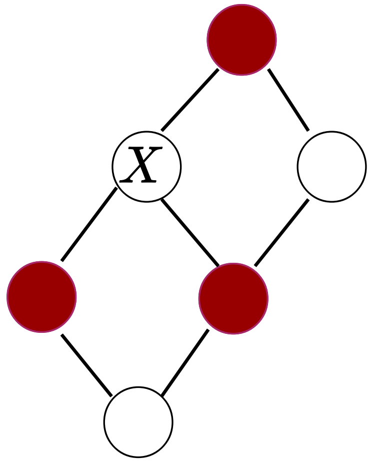
$$X \perp Y \mid \mathcal{X} - \{X, Y\}$$



Local conditional Independencies

$MB^{\mathcal{H}}(X)$: **Markov blanket** of node X in graph *H*

$$X \perp \mathcal{X} - X - MB^{\mathcal{H}}(X) \mid MB^{\mathcal{H}}$$

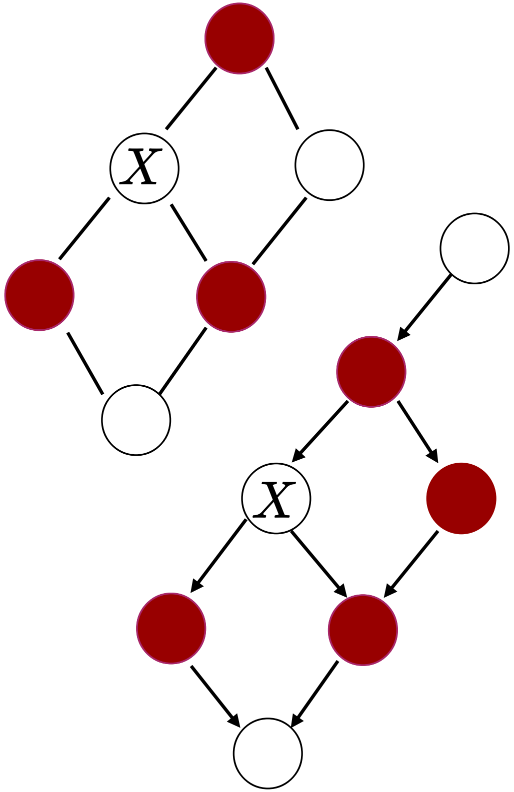


Given its Markov blanket X is independent of every other variable

Local conditional Independencies

$MB^{\mathcal{H}}(X)$: **Markov blanket** of X in graph *H*

$$X \perp \mathcal{X} - X - MB^{\mathcal{H}}(X) \mid MB^{\mathcal{H}}$$



$MB^{\mathcal{G}}(X)$: **Markov blanket** of X in DAG *G*

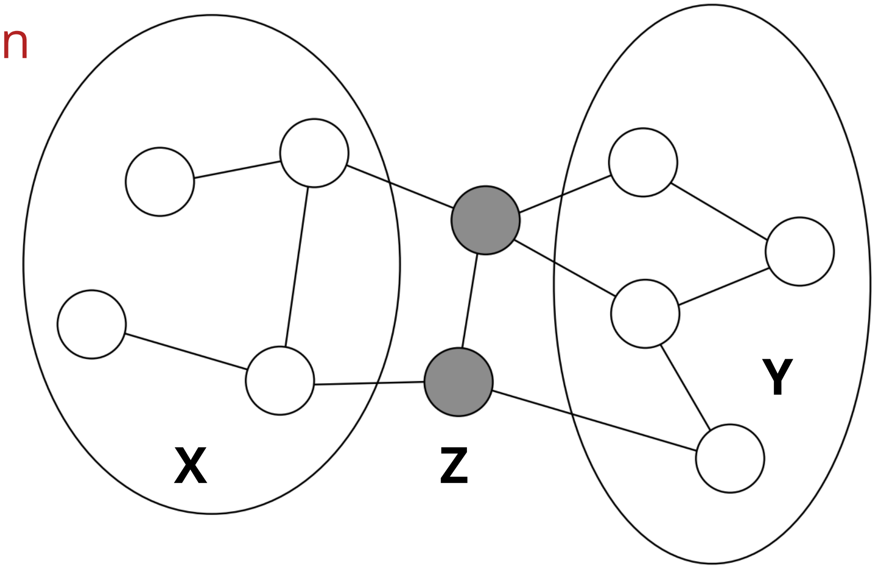
- Parents
- *Children*
- *Parents of children*

$$X \perp \mathcal{X} - X - MB^{\mathcal{G}}(X) \mid MB^{\mathcal{G}}$$

Global conditional Independencies

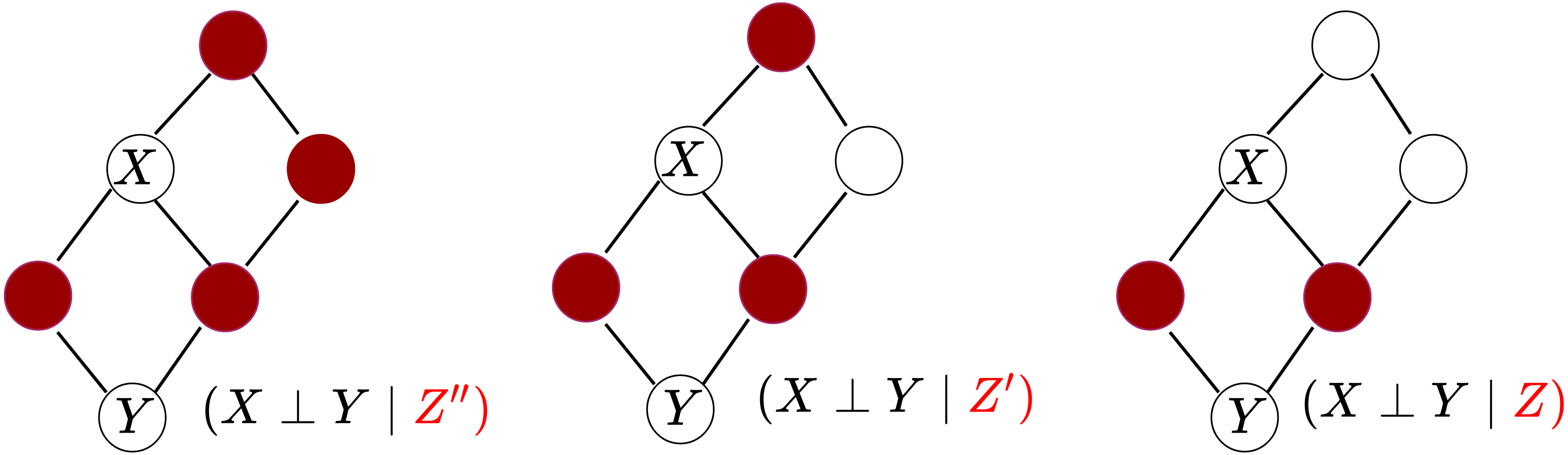
$X \perp Y \mid Z$ iff every path between **X** and **Y** is blocked by **Z**

much simpler than D-separation



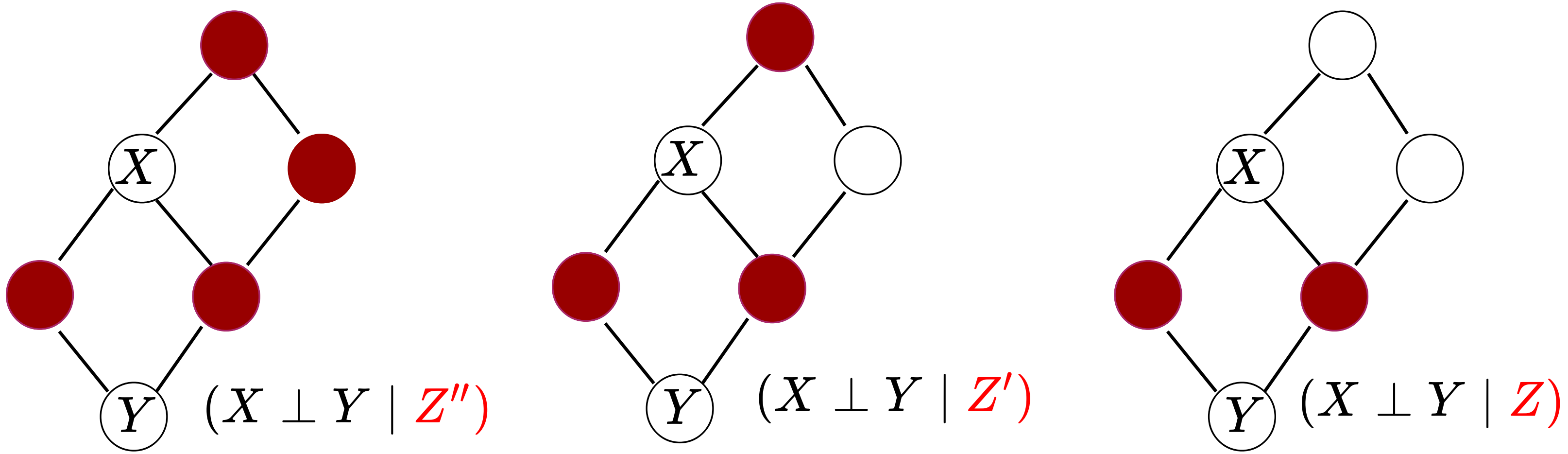
Relationship between the three

pairwise $\mathcal{I}_p$ $\Leftarrow$ local $\mathcal{I}_\ell$ $\Leftarrow$ global $\mathcal{I}$



Relationship between the three

pairwise $\mathcal{I}_p$ $\Leftarrow$ local $\mathcal{I}_\ell$ $\Leftarrow$ global $\mathcal{I}$



P>0: pairwise $\mathcal{I}_p$ $\Rightarrow$ local $\mathcal{I}_\ell$ $\Rightarrow$ global $\mathcal{I}$

Factorization & independence

Recall this relationship in **Bayesian Networks**:

- **Factorization** according to a DAG
- **Local** & **global** CIs

Equivalent

(same family of distributions)

Is it similar for **Markov Networks**?

- **Factorization** according to an *undirected graph*
- **Pairwise, local & global** CIs

Equivalent?

Factorization & Independence

Is it similar for **Markov Networks**?

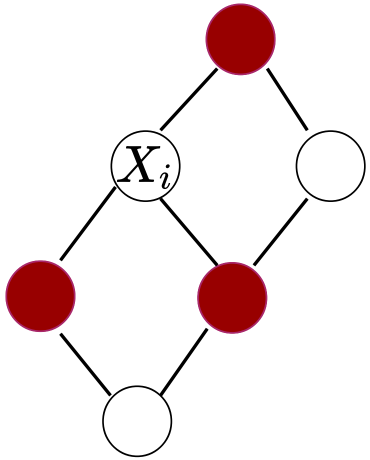
- **Factorization** according to an *undirected graph*
- **Pairwise, local & global** CIs

Short answer:

- for positive distributions they are equivalent

Factorization $\Rightarrow$ CI

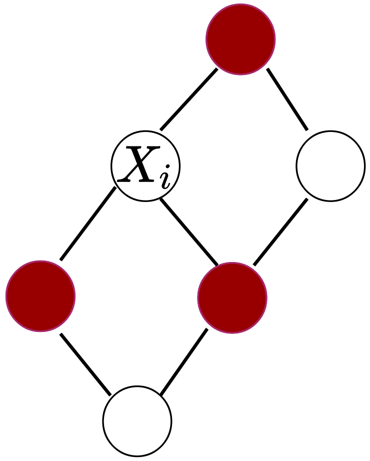
given $P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{C}_k)$ does **local** CI hold?



Factorization $\Rightarrow$ CI

given $P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{C}_k)$ does **local** CI hold?

proof



Factorization $\Rightarrow$ CI

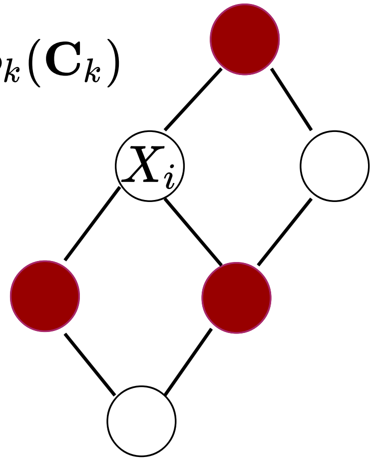
given $P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{C}_k)$ does **local** CI hold?

proof

$P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{C}_k) = \prod_{\mathbf{C}_k \in MB(X_i)} \phi_k(\mathbf{C}_k) \prod_{\mathbf{C}_k \notin MB(X_i)} \phi_k(\mathbf{C}_k)$

$= f(X_i, MB(X_i))\, g(\mathcal{X} - X_i) \quad \Rightarrow$

$X_i \perp \mathcal{X} - MB^{\mathcal{H}}(X_i) - X_i \mid MB^{\mathcal{H}}(X_i)$



CI $\Rightarrow$ factorization

Hammersely-Clifford theorem:

If P is *strictly positive* satisfying CI $\mathcal{I}(\mathcal{H})$
then P factorizes over $\mathcal{H}$

proof

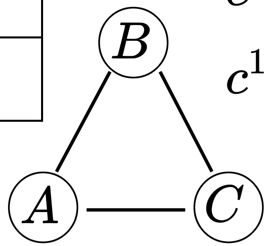
needs canonical parametrization 

Parametrization: redundancy

is this representation of P unique?

$$P(A,B,C) \propto \phi_1(A,B)\phi_2(B,C)\phi_3(C,A)$$

	b^0	b^1
a^0	1	2
a^1	2	1



	b^0	b^1
c^0	1	1
c^1	2	2

	c^0	c^1
a^0	4	8
a^1	16	1

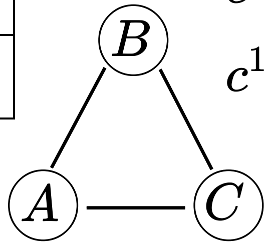
Parametrization: redundancy

is this representation of P unique?

$$P(A, B, C) = \frac{1}{Z}\phi_1(A, B)\phi_2(B, C)\phi_3(C, A)$$

multiplying all factors by a constant only affects **Z**

	b^0	b^1
a^0	10	20
a^1	20	10



	b^0	b^1
c^0	10	10
c^1	20	20

	c^0	c^1
a^0	.4	.8
a^1	1.6	.1

Parametrization: redundancy

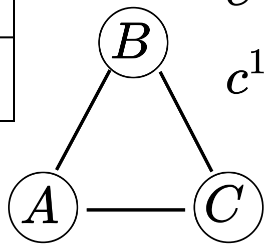
is this representation of P unique?

$$P(A,B,C) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\phi_3(C,A)$$

use the logarithmic form

$$P(A,B,C) = \frac{1}{Z}2^{(\psi_1(A,B)+\psi_2(B,C)+\psi_3(C,A))}$$

	b^0	b^1
a^0	0	1
a^1	1	0



log-values

	b^0	b^1
c^0	0	0
c^1	1	1

	c^0	c^1
a^0	2	3
a^1	4	0

Parametrization: redundancy

Is this representation of P unique?

$$P(A,B,C) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\phi_3(C,A)\phi_4(B)\phi_5(A)\phi_6(C)$$

use the logarithmic form

$$P(A,B,C) = \frac{1}{Z}2^{(\psi_1(A,B)+\psi_2(B,C)+\psi_3(C,A))}$$

simplify using local potentials

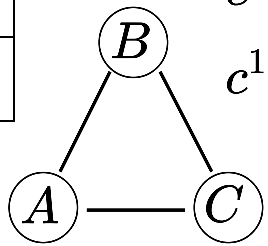
$$P(A,B,C) = \frac{1}{Z}2^{(\psi_1(A,B)+\psi_2(B,C)+\psi_3(C,A)+\psi_4(A)+\psi_5(B)+\psi_6(C))}$$

log-values

b^0	b^1
0	0

	b^0	b^1
a^0	0	1
a^1	1	0

	b^0	b^1
c^0	0	0
c^1	0	0



a^0	a^1
2	0

c^0	c^1
0	1

	c^0	c^1
a^0	0	1
a^1	4	0

Parametrization: redundancy

is this representation of P unique?

$$P(A,B,C) = \frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\phi_3(C,A)\phi_5(A)\phi_6(C)$$

use the logarithmic form

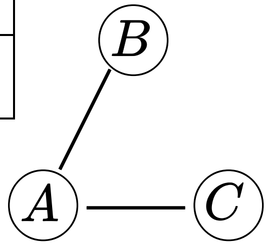
$$P(A,B,C) = \frac{1}{Z}2^{(\psi_1(A,B)+\psi_2(B,C)+\psi_3(C,A))}$$

simplify using local potentials

$$P(A,B,C) = \frac{1}{Z}2^{(\psi_1(A,B)+\psi_2(B,C)+\psi_3(C,A)+\psi_4(A)+\psi_5(B)+\psi_6(C))}$$

log-values

	b^0	b^1
a^0	0	1
a^1	1	0



c^0	c^1
0	1

a^0	a^1
2	0

	c^0	c^1
a^0	0	1
a^1	4	0

Parametrization: example (Ising model)

Ising model: $Val(X_i) = \{-1, +1\}$ $p(\mathbf{x}) = \frac{1}{Z(t)} \exp\left(-\frac{1}{t}\left(\sum_i h_i x_i + \frac{1}{2}\sum_{i,j \in \mathcal{E}} x_i J_{ij} x_j\right)\right)$

can represent all positive, pairwise Markov networks over the binary domain

interactions

	-1	$+1$
-1	J	-J
$+1$	-J	J

local *field*

-1	$+1$
-h	h

log-values

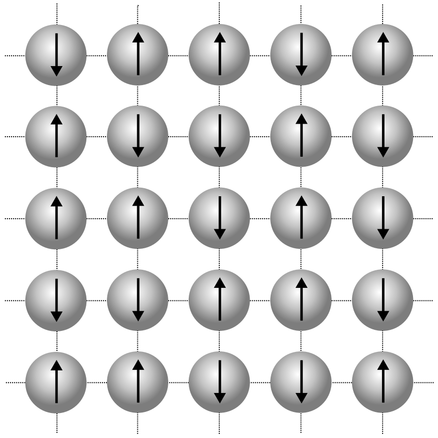


Image: https://web.stanford.edu/~peastman/statmech/phasetransitions.html

Parametrization: example **(Boltzmann machine)**

Boltzmann machine: $Val(X_i) = \{0, 1\}$

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_i b_i x_i - \frac{1}{2}\sum_{i,j \in \mathcal{E}} x_i W_{ij} x_j\right)$$

interaction weights

	-1	$+1$
-1	0	0
$+1$	0	J

local *bias*

-1	$+1$
0	h

log-values

Parametrization: log-linear model

for a positive distribution:

$$P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{D}_k) = \exp(-\sum_k \psi_k(\mathbf{D}_k))$$

energy $-\log(\phi_k(\mathbf{D}_k))$

Parametrization: log-linear model

for a positive distribution:

$$P(\mathbf{X}) \propto \prod_k \phi_k(\mathbf{D}_k) = \exp(-\sum_k \underline{\psi_k(\mathbf{D}_k)})$$

energy $-\log(\phi_k(\mathbf{D}_k))$

linearly parameterize it:

$$P_w(\mathbf{X}) \propto \exp(-\sum_k w_k \underline{f_k(\mathbf{D}_k)})$$

feature/sufficient statistics

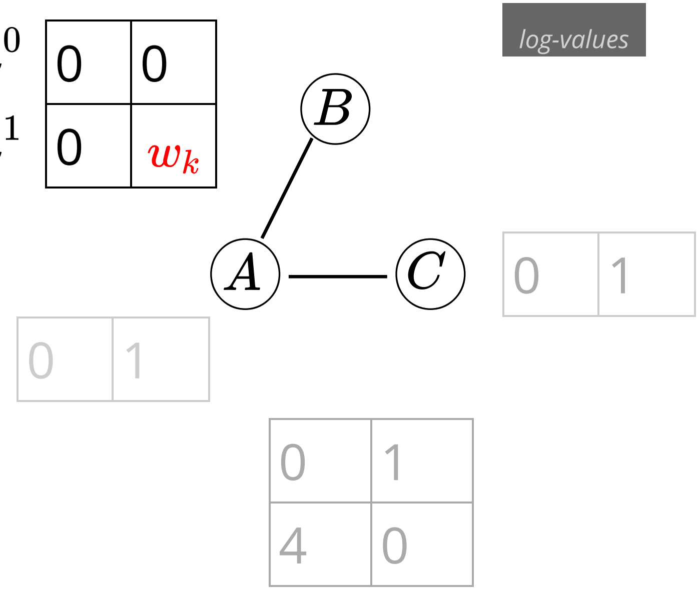
Parametrization: log-linear model

features in **discrete** distributions:

$$P_w(\mathbf{X}) \propto \exp\left(-\sum_k w_k f_k(\mathbf{D}_k)\right)$$

$$f_{1,1}(A, B) = \mathbb{I}(A = a^1, B = b^1)$$

	b^0	b^1
a^0	0	0
a^1	0	w_k



Parametrization: log-linear model

$$P_w(\mathbf{X}) \propto \exp(-\sum_k w_k f_k(\mathbf{D}_k))$$

$\downarrow$

$$f_{1,1}(A,B) = \mathbb{I}(A = a^1, B = b^1)$$

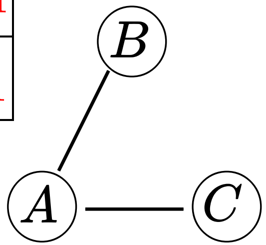
$$f_{0,1}(A,B) = \mathbb{I}(A = a^0, B = b^1)$$

$$f_{1,0}(A,B) = \mathbb{I}(A = a^1, B = b^0)$$

$$f_{0,0}(A,B) = \mathbb{I}(A = a^0, B = b^0)$$

	b^0	b^1
a^0	$w_{0,0}$	$w_{0,1}$
a^1	$w_{1,0}$	$w_{1,1}$

log-values



Parametrization: log-linear model

$$P_w(\mathbf{X}) \propto \exp(-\textstyle\sum_k w_k f_k(\mathbf{D}_k))$$

$\downarrow$

$$f_{1,1}(A,B) = \mathbb{I}(A = a^1, B = b^1)$$

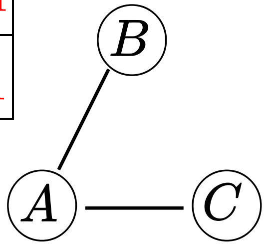
$$f_{0,1}(A,B) = \mathbb{I}(A = a^0, B = b^1)$$

$$f_{1,0}(A,B) = \mathbb{I}(A = a^1, B = b^0)$$

$$f_{0,0}(A,B) = \mathbb{I}(A = a^0, B = b^0)$$

	b^0	b^1
a^0	$w_{0,0}$	$w_{0,1}$
a^1	$w_{1,0}$	$w_{1,1}$

log-values



Overparameterized model: $\{w_k\} \rightarrow P_w$ is not one-to-one

Parametrization: log-linear model

$$P_w(\mathbf{X}) \propto \exp(-\sum_k w_k f_k(\mathbf{D}_k))$$

Redundant $\equiv$ linearly dependent features

$$\sum_k \alpha_k f_k(\mathbf{D}) = \alpha \quad \forall \mathbf{D}$$

$\downarrow$

$$P_w(\mathbf{X}) \propto \exp(-\sum_k w_k f_k(\mathbf{D}_k)) \propto exp(-\sum_k (w_k + \alpha_k) f_k(\mathbf{D_k})) \propto P_{w+\alpha}(\mathbf{X})$$

Parametrization: log-linear model

$$P_w(\mathbf{X}) \propto \exp(-\textstyle\sum_k w_k f_k(\mathbf{D}_k))$$

$$\downarrow$$

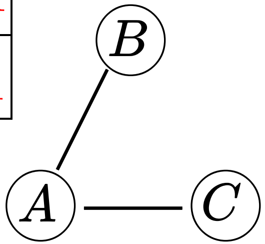
$$f_{1,1}(A,B) = \mathbb{I}(A = a^1, B = b^1)$$

$$f_{0,1}(A,B) = \mathbb{I}(A = a^0, B = b^1)$$

$$f_{1,0}(A,B) = \mathbb{I}(A = a^1, B = b^0)$$

$$f_{0,0}(A,B) = \mathbb{I}(A = a^0, B = b^0)$$

	b^0	b^1
a^0	$w_{0,0}$	$w_{0,1}$
a^1	$w_{1,0}$	$w_{1,1}$



log-values

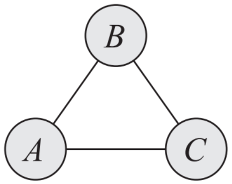
Linear dependency of features:

$$f_{0,0}(A,B) + f_{1,0}(A,B) + f_{0,1}(A,B) + f_{1,1}(A,B) = 1$$

Parametrization: factor-graph

Markov network representation:

- identifies CI
- defines the factorized form
  - is not fine-grained enough



$$P(A, B, C) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, A) \quad ?$$

$$P(A, B, C) = \phi_1(A, B, C) \quad ?$$

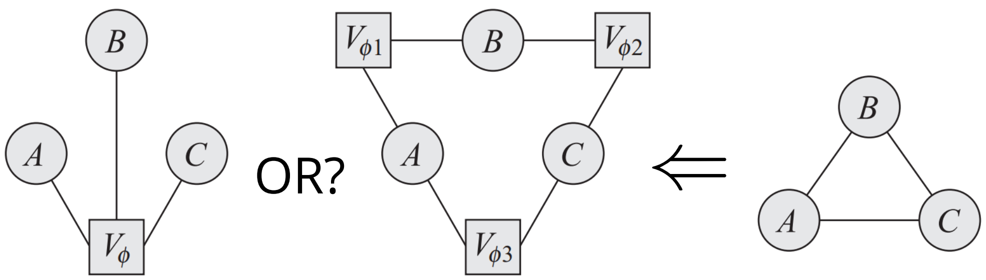
Parametrization: factor-graph

use a bipartite structure:

- factors (square)
- variables (circle)

$$P(A, B, C) = \phi_1(A, B)\phi_2(B, C)\phi_3(C, A)$$

$$P(A, B, C) = \phi(A, B, C)$$



Summary

- similar to directed models:
  - factorization of the probability over cliques
  - set of conditional independencies
    - (pariwise, local, global)

$P > 0 \Rightarrow$

same family of dists.

Summary

- similar to directed models:
  - factorization of the probability over cliques
  - set of conditional independencies
    - (pariwise, local, global)

$P > 0 \Rightarrow$

same family of dists.

- parametrization
  - redundancy (same dist. different params/factors)
  - log-linear model
  - factor-graph (finer-grained specification of the factors)

Bonus Slides

Parametrization: canonical form

reparameterize a given Gibbs dist.

$$P(\mathbf{X}) \propto \exp(-\sum_k \psi_k(\mathbf{D}_k))$$

such that low order interactions are automatically moved to smaller cliques

need to fixed an assignment $\xi^* = (x_1^*, \ldots, x_n^*) \quad e.g., \quad \xi^* = (0, \ldots, 0)$

Mobius inversion lemma

For two functions $f, g : 2^{\mathcal{X}} \rightarrow \Re$ defined over all subsets $\mathcal{Z} \subseteq \mathcal{X}$ the following are equivalent:

$$\forall \mathcal{Z} \subseteq \mathcal{X} \quad f(\mathcal{Z}) = \sum_{\mathcal{S} \subseteq \mathcal{Z}} g(\mathcal{S})$$

$$\forall \mathcal{Z} \subseteq \mathcal{X} \quad g(\mathcal{Z}) = \sum_{\mathcal{S} \subseteq \mathcal{Z}} (-1)^{|\mathcal{Z} - \mathcal{S}|} f(\mathcal{S})$$

Mobius inversion

Parametrization: canonical form

Given a *fixed an assignment* $\xi^* = (x_1^*, \ldots, x_n^*)$ $e.g.,$ $\xi^* = (0, \ldots, 0)$

$f(x_{\mathcal{Z}}) \triangleq \log P(x_{\mathcal{Z}}, \xi^*_{-\mathcal{Z}})$ is defined for all $\mathcal{Z} \subseteq \{1, \ldots, N\}$

$$f(x) = \log P(x)$$

Parametrization: canonical form

Given a *fixed an assignment* $\xi^* = (x_1^*, \ldots, x_n^*)$ *e.g.,* $\xi^* = (0, \ldots, 0)$

$$f(x_{\mathcal{Z}}) \triangleq \log P(x_{\mathcal{Z}}, \xi^*_{-\mathcal{Z}})$$ is defined for all $\mathcal{Z} \subseteq \{1, \ldots, N\}$

$$f(x) = \log P(x)$$

Its *Mobius inversion*: $g(x_{\mathcal{Z}}) = -\sum_{\mathcal{S} \subseteq \mathcal{Z}} (-1)^{|\mathcal{Z} - \mathcal{S}|} \log P(x_{\mathcal{S}}, \xi^*_{-\mathcal{S}})$

Parametrization: canonical form

Given a *fixed an assignment* $\xi^* = (x_1^*, \ldots, x_n^*)$ *e.g.,* $\xi^* = (0, \ldots, 0)$

$f(x_{\mathcal{Z}}) \triangleq \log P(x_{\mathcal{Z}}, \xi^*_{-\mathcal{Z}})$ is defined for all $\mathcal{Z} \subseteq \{1, \ldots, N\}$

$f(x) = \log P(x)$

Its *Mobius inversion*: $g(x_{\mathcal{Z}}) = -\sum_{\mathcal{S} \subseteq \mathcal{Z}} (-1)^{|\mathcal{Z}-\mathcal{S}|} \log P(x_{\mathcal{S}}, \xi^*_{-\mathcal{S}})$

Define **factors** over each subset of nodes: $\psi_{\mathcal{Z}}(x_{\mathcal{Z}}) = -g(x_{\mathcal{Z}})$

Parametrization: canonical form

Given a *fixed an assignment* $\xi^* = (x_1^*, \ldots, x_n^*)$ $e.g.,$ $\xi^* = (0, \ldots, 0)$

$f(x_\mathcal{Z}) \triangleq \log P(x_\mathcal{Z}, \xi^*_{-\mathcal{Z}})$ is defined for all $\mathcal{Z} \subseteq \{1, \ldots, N\}$

$f(x) = \log P(x)$

Its *Mobius inversion*: $g(x_\mathcal{Z}) = -\sum_{\mathcal{S} \subseteq \mathcal{Z}} (-1)^{|\mathcal{Z} - \mathcal{S}|} \log P(x_\mathcal{S}, \xi^*_{-\mathcal{S}})$

Define **factors** over each subset of nodes: $\psi_\mathcal{Z}(x_\mathcal{Z}) = -g(x_\mathcal{Z})$

From Mobius lemma: $P(x) = \exp(-\sum \psi_\mathcal{Z}(x_\mathcal{Z}))$

Parametrization: canonical form

Its *Mobius inversion*: $g(x_{\mathcal{Z}}) = -\sum_{\mathcal{S}\subseteq\mathcal{Z}}(-1)^{|\mathbf{Z}-\mathbf{S}|}\log P(x_{\mathcal{S}}, \xi^*_{-\mathcal{S}})$

Define **factors** over each subset of nodes: $\psi_{\mathcal{Z}}(x_{\mathcal{Z}}) = -g(x_{\mathcal{Z}})$

From Mobius lemma: $P(x) = \exp(-\sum \psi_{\mathcal{Z}}(x_{\mathcal{Z}}))$

Problem: one factor per subset of nodes

Proof of Hammersly-Clifford theorem:

When $\mathcal{Z}$ is not a clique $\psi_{\mathcal{Z}}(x_{\mathcal{Z}})$ becomes zero.

Proof of the Hammersley-Clifford

Recap:

- fix an assignment
- define factors over each subset of nodes as:

$$\psi(x_{\mathcal{Z}}) = \sum_{\mathcal{S} \subseteq \mathcal{Z}} (-1)^{|\mathcal{Z} - \mathcal{S}|} \log P(x_{\mathcal{S}}, \xi^*_{-\mathcal{S}})$$

- if $\mathcal{Z}$ is not a clique in $\mathcal{H}$ then $\exists_{i,j \in \mathcal{Z}} X_i \perp X_j \mid \mathbf{X} - \{X_i, X_j\}$
  - we can show that $\psi_{\mathcal{Z}}(x_{\mathcal{Z}}) = 0 \quad \forall x_{\mathcal{Z}}$