# Stat 375: Inference in Graphical Models
# Lectures 7-8-9

Andrea Montanari

Stanford University

May 6, 2012

# Variational methods

**Idea**

I know a lot about (convex) optimization. . .

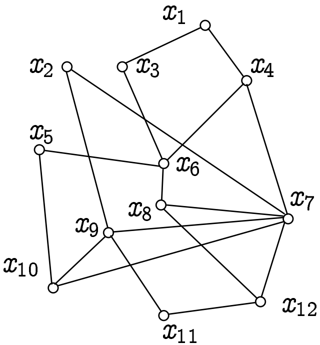
# Variational methods

## Idea

I know a lot about (convex) optimization. . .

# Outline

# Undirected Pairwise Graphical Model




$G = (V, E)$, $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \mathcal{X}$, $|\mathcal{X}| < \infty$

$$\mu(x) = \frac{1}{Z} \prod_{(ij) \in E} \psi_{ij}(x_i, x_j)\,.$$

# Notation

$$\text{`Actual' probability} \longrightarrow \mu$$

$$\text{`Trial' probability (`belief')} \longrightarrow b$$

# Gibbs Variational Principle

# I want to compute

$$\Phi \equiv \log Z$$

# Gibbs Free Energy

$$\mathbb{G}_\psi : \quad \mathcal{X}^V \to \mathbb{R}\,,$$

$$b \mapsto \mathbb{G}_\psi(b)\,.$$

**Proposition**

*$\mathbb{G}_\psi$ is strictly concave, and achieves its unique maximum at $b = \mu$. Further*

$$\mathbb{G}_\psi(b = \mu) = \Phi\,.$$

# Gibbs Free Energy

$$\mathbb{G}_\psi : \qquad \mathcal{X}^V \to \mathbb{R}\,,$$

$$b \mapsto \mathbb{G}_\psi(b)\,.$$

## Proposition

*$\mathbb{G}_\psi$ is strictly concave, and achieves its unique maximum at $b = \mu$. Further*

$$\mathbb{G}_\psi(b = \mu) = \Phi\,.$$

# Gibbs Free Energy

$$\mu(x) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j) \equiv \frac{1}{Z}\, \psi_{\rm tot}(x)\,.$$

Definition

$$\begin{aligned} \mathbb{G}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H(b) \\ &= \sum_{(ij)\in E} \sum_{x_i, x_j \in \mathcal{X}} b(x_i, x_j) \log \psi_{ij}(x_i, x_j) - \sum_{x\in\mathcal{X}^V} b(x) \log b(x) \end{aligned}$$

# Gibbs Free Energy

$$\mu(x) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j) \equiv \frac{1}{Z}\, \psi_{\rm tot}(x)\,.$$

Definition

$$\begin{aligned} \mathbb{G}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H(b) \\ &= \sum_{(ij)\in E} \sum_{x_i, x_j \in \mathcal{X}} b(x_i, x_j) \log \psi_{ij}(x_i, x_j) - \sum_{x\in\mathcal{X}^V} b(x) \log b(x) \end{aligned}$$

# Gibbs Free Energy

$$\mu(x) = \frac{1}{Z} \prod_{(ij)\in E} \psi_{ij}(x_i, x_j) \equiv \frac{1}{Z}\, \psi_{\rm tot}(x)\, .$$

Definition

$$\begin{aligned} \mathbb{G}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H(b) \\ &= \sum_{(ij)\in E} \sum_{x_i, x_j \in \mathcal{X}} b(x_i, x_j) \log \psi_{ij}(x_i, x_j) - \sum_{x\in\mathcal{X}^V} b(x) \log b(x) \end{aligned}$$

# Proof

1. Define the Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{G}(b) - \lambda\Big\{ \sum_{x\in\mathcal{X}^V} b(x) - 1 \Big\}$$

and differentiate

2. Observe that

$$z \mapsto z\log z \quad \text{is convex on } \mathbb{R}_+$$

# Proof

**1.** Define the Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{G}(b) - \lambda\Big\{\sum_{x\in\mathcal{X}^V} b(x) - 1\Big\}$$

and differentiate

**2.** Observe that

$$z \mapsto z\log z \quad \text{is convex on } \mathbb{R}_+$$

# Naive Mean Field

# Good news/Bad news

Counting = Convex Optimization

$\mathrm{M}(\mathcal{X}^V)$ is $|\mathcal{X}|^V - 1$ dimensional.

# Good news/Bad news

Counting = Convex Optimization

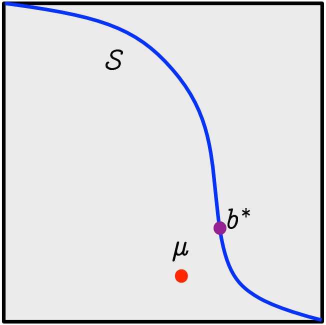
$\mathrm{M}(\mathcal{X}^V)$ is $|\mathcal{X}|^V - 1$ dimensional.

# Good news/Bad news

Counting = Convex Optimization

$\mathsf{M}(\mathcal{X}^V)$ is $|\mathcal{X}|^V - 1$ dimensional.

# Idea




Maximize Gibbs free energy on a low-dim subset

$$\Phi \geq \sup_{b \in \mathcal{S}} \mathbb{G}_\psi(b) .$$

# Naive Mean Field Idea

$$\mathcal{S} = \left\{ b \in \mathsf{M}(\mathcal{X}^n) : \quad b = b_1 \times b_2 \times \cdots \times b_n \right\},$$

$$\text{Abuse} \qquad b \equiv \{b_i\}_{i \in V}$$

$$\mathbb{F}_{\rm MF}(b = \{b_i\}_{i \in V}) = \mathbb{G}_\psi(b_1 \times \cdots \times b_n)$$

# Naive Mean Field Idea

$$\mathcal{S} = \left\{ b \in \mathsf{M}(\mathcal{X}^n) : \quad b = b_1 \times b_2 \times \cdots \times b_n \right\},$$

Abuse $\qquad b \equiv \{b_i\}_{i \in V}$

$$\mathbb{F}_{\rm MF}(b = \{b_i\}_{i \in V}) = \mathbb{G}_\psi(b_1 \times \cdots \times b_n)$$

# Naive Mean Field Idea

$$\mathcal{S} = \Big\{ b \in \mathsf{M}(\mathcal{X}^n) : \quad b = b_1 \times b_2 \times \cdots \times b_n \Big\},$$

$$\text{Abuse} \qquad b \equiv \{b_i\}_{i \in V}$$

$$\mathbb{F}_{\mathrm{MF}}(b = \{b_i\}_{i \in V}) = \mathbb{G}_\psi(b_1 \times \cdots \times b_n)$$

# Explicitly

$$
\begin{aligned}
\mathbb{F}_{\rm MF}(b) &= \sum_{(i,j)\in E} \mathbb{E}_{b_i\times b_j} \log\psi_{ij}(x_i,x_j) + H(b_1\times\cdots\times b_n)\\
&= \sum_{(i,j)\in E}\sum_{x_i,x_j} b_i(x_i)b_j(x_j)\log\psi_{ij}(x_i,x_j) - \sum_{x_i} b_i(x_i)\log b_i(x_i)
\end{aligned}
$$

**Problem:** Not convex.

# Explicitly

$$
\begin{aligned}
\mathbb{F}_{\mathrm{MF}}(b) &= \sum_{(i,j)\in E} \mathbb{E}_{b_i \times b_j} \log \psi_{ij}(x_i, x_j) + H(b_1 \times \cdots \times b_n) \\
&= \sum_{(i,j)\in E} \sum_{x_i, x_j} b_i(x_i) b_j(x_j) \log \psi_{ij}(x_i, x_j) - \sum_{x_i} b_i(x_i) \log b_i(x_i)
\end{aligned}
$$

**Problem:** Not convex.

# Explicitly

$$\begin{aligned}\mathbb{F}_{\rm MF}(b) &= \sum_{(i,j)\in E} \mathbb{E}_{b_i\times b_j} \log\psi_{ij}(x_i,x_j) + H(b_1\times\cdots\times b_n)\\ &= \sum_{(i,j)\in E}\sum_{x_i,x_j} b_i(x_i)b_j(x_j)\log\psi_{ij}(x_i,x_j) - \sum_{x_i} b_i(x_i)\log b_i(x_i)\end{aligned}$$

**Problem:** Not convex.

# Stationarity condition

Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{F}_{\rm MF}(b) + \sum_{i\in V}\lambda_i\Big\{\sum_{x_i\in\mathcal{X}} b_i(x_i) - 1\Big\},$$

$$\frac{\partial\mathcal{L}}{\partial b_i(x_i)}(b,\lambda) = \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0,$$

$$b_i(x_i) \cong \exp\Big\{\sum_{j\in\partial i}\sum_{x_j}\log\psi_{ij}(x_i,x_j)b_j(x_j)\Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i),$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Stationarity condition

Lagrangian

$$
\mathcal{L}(b,\lambda) = \mathbb{F}_{\rm MF}(b) + \sum_{i\in V} \lambda_i \Big\{ \sum_{x_i\in\mathcal{X}} b_i(x_i) - 1 \Big\},
$$

$$
\frac{\partial \mathcal{L}}{\partial b_i(x_i)}(b,\lambda) = \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0,
$$

$$
b_i(x_i) \cong \exp\Big\{ \sum_{j\in\partial i}\sum_{x_j} \log\psi_{ij}(x_i,x_j) b_j(x_j) \Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i),
$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Stationarity condition

Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{F}_{\rm MF}(b) + \sum_{i\in V} \lambda_i \Big\{ \sum_{x_i\in\mathcal{X}} b_i(x_i) - 1 \Big\},$$

$$\frac{\partial \mathcal{L}}{\partial b_i(x_i)}(b,\lambda) = \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0,$$

$$b_i(x_i) \cong \exp\Big\{ \sum_{j\in\partial i}\sum_{x_j} \log\psi_{ij}(x_i,x_j) b_j(x_j) \Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i),$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Stationarity condition

Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{F}_{\rm MF}(b) + \sum_{i\in V}\lambda_i\Big\{\sum_{x_i\in\mathcal{X}} b_i(x_i) - 1\Big\},$$

$$\frac{\partial \mathcal{L}}{\partial b_i(x_i)}(b,\lambda) = \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0,$$

$$b_i(x_i) \cong \exp\Big\{\sum_{j\in\partial i}\sum_{x_j}\log\psi_{ij}(x_i,x_j)b_j(x_j)\Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i),$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Stationarity condition

Lagrangian

$$\mathcal{L}(b,\lambda) = \mathbb{F}_{\rm MF}(b) + \sum_{i\in V} \lambda_i \Big\{ \sum_{x_i\in\mathcal{X}} b_i(x_i) - 1 \Big\},$$

$$\frac{\partial \mathcal{L}}{\partial b_i(x_i)}(b,\lambda) = \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0,$$

$$b_i(x_i) \cong \exp\Big\{ \sum_{j\in\partial i}\sum_{x_j} \log\psi_{ij}(x_i,x_j) b_j(x_j)\Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i),$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Stationarity condition

Lagrangian

$$\begin{aligned}
\mathcal{L}(b,\lambda) &= \mathbb{F}_{\rm MF}(b) + \sum_{i\in V}\lambda_i\Big\{\sum_{x_i\in\mathcal{X}} b_i(x_i) - 1\Big\}\,,\\
\frac{\partial\mathcal{L}}{\partial b_i(x_i)}(b,\lambda) &= \sum_{j\in\partial i}\sum_{x_j\in\mathcal{X}} b_j(x_j)\log\psi_{ij}(x_i,x_j) - 1 - \log b_i(x_i) + \lambda_i = 0\,,\\
b_i(x_i) &\cong \exp\Big\{\sum_{j\in\partial i}\sum_{x_j}\log\psi_{ij}(x_i,x_j)b_j(x_j)\Big\} \equiv \mathsf{F}_{\rm MF}(b)_i(x_i)\,,
\end{aligned}$$

[Naive Mean Field Equations]

Typically a fixed point is searched by iteration: $b^{(t+1)} = \mathsf{F}_{\rm MF}(b^{(t)})$

# Another point of view

**Exercise :** For $\psi_{ij}(x_i, x_j) = e^{\theta_{ij}(x_i,x_j)}$

$$\mu_i(x_i) \;\; \cong \;\; \mathbb{E}_\mu \left\{ \frac{e^{\sum_{j\in\partial i} \theta_{ij}(x_i,X_j)}}{\sum_{x_i'} e^{\sum_{j\in\partial i} \theta_{ij}(x_i',X_j)}} \right\}$$

If we could move the expectation to exponents

$$\mu_i(x_i) \;\; \cong \;\; e^{\mathbb{E}_\mu \sum_{j\in\partial i} \theta_{ij}(x_i,X_j)}$$

[= Naive Mean Field]

# Another point of view

**Exercise :** For $\psi_{ij}(x_i, x_j) = e^{\theta_{ij}(x_i,x_j)}$

$$\mu_i(x_i) \;\cong\; \mathbb{E}_\mu \left\{ \frac{e^{\sum_{j\in\partial i}\theta_{ij}(x_i,X_j)}}{\sum_{x_i'} e^{\sum_{j\in\partial i}\theta_{ij}(x_i',X_j)}} \right\}$$

If we could move the expectation to exponents

$$\mu_i(x_i) \;\cong\; e^{\mathbb{E}_\mu \sum_{j\in\partial i}\theta_{ij}(x_i,X_j)}$$

[= Naive Mean Field]

# Bethe Free Energy

# Problem

One dimensional marginals give a very poor approximation.

**Example:** $x_1, x_2 \in \{0, 1\}$

$$\mu(x) = \frac{1}{2}\,\mathbb{I}(x_1 \oplus x_2 = 0)$$

Would like to account exactly for the correlations induced by edges.

# Problem

One dimensional marginals give a very poor approximation.

**Example:** $x_1, x_2 \in \{0, 1\}$

$$\mu(x) = \frac{1}{2}\,\mathbb{I}(x_1 \oplus x_2 = 0)$$

Would like to account exactly for the correlations induced by edges.

# Would like

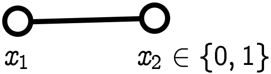
$$\begin{aligned} \mathbb{F}: \quad \mathsf{M}(\mathcal{X}\times\mathcal{X})^E \times \mathsf{M}(\mathcal{X})^V &\to \mathbb{R} \\ b = \{b_{ij}, b_i\}_{(i,j)\in E, i\in V} &\mapsto \mathbb{F}(b) \end{aligned}$$

$$b_{ij} = b_{ij}(x_i, x_j),\; b_i = b_i(x_i),$$

such that

$$\begin{aligned} \arg\max_b \mathbb{F}(b) &\approx \mu\,, \\ \max_b \mathbb{F}(b) &\approx \Phi\,, \end{aligned}$$

# What is really the domain?




**Example:**

$$b_1 = \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix}, \qquad b_2 = \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix}, \qquad b_1 = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}.$$

# Natural guess

$$b \in \mathrm{MARG}(G)$$

$$\mathrm{MARG}(G) = \Big\{ b = \{b_i, b_{ij}\} : \text{ marginals of a distribution on } \mathcal{X}^V \Big\},$$

$$b_i(x_i) = \sum_{x_{V \setminus i}} p(x),$$

$$b_{i,j}(x_i, x_j) = \sum_{x_{V \setminus \{i,j\}}} p(x),$$

$$p \in \mathsf{M}(\mathcal{X}^V).$$

# Natural guess

$$b \in \mathrm{MARG}(G)$$

$$\mathrm{MARG}(G) = \Big\{ b = \{b_i, b_{ij}\} : \text{ marginals of a distribution on } \mathcal{X}^V \Big\},$$

$$b_i(x_i) = \sum_{x_{V\setminus i}} p(x),$$

$$b_{i,j}(x_i, x_j) = \sum_{x_{V\setminus \{i,j\}}} p(x),$$

$$p \in \mathsf{M}(\mathcal{X}^V).$$

# Natural guess

$$b \in \mathrm{MARG}(G)$$

$$\mathrm{MARG}(G) = \Big\{ b = \{b_i, b_{ij}\} : \text{ marginals of a distribution on } \mathcal{X}^V \Big\},$$

$$b_i(x_i) = \sum_{x_{V\setminus i}} p(x)\,,$$

$$b_{i,j}(x_i, x_j) = \sum_{x_{V\setminus\{i,j\}}} p(x)\,,$$

$$p \in \mathsf{M}(\mathcal{X}^V)\,.$$

# Bad news

In general checking $b \in \text{MARG}(G)$ in NP-hard.

# Second attempt

$$b \in \mathrm{LOC}(G)$$

$$\mathrm{LOC}(G) = \Big\{ b = \{b_i, b_{ij}\} : \ \text{locally consistent marginals} \Big\},$$

$$\Big\{ \{b_i, b_{ij}\} : \quad \sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i), \quad \sum_{x_i} b_i(x_i) = 1 \quad \Big\}$$

# Second attempt

$$b \in \mathrm{LOC}(G)$$

$$\mathrm{LOC}(G) = \Big\{ b = \{b_i, b_{ij}\} : \text{ locally consistent marginals} \Big\},$$

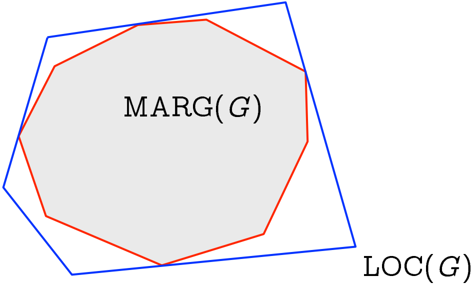
$$\Big\{ \{b_i, b_{ij}\} : \quad \sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i), \quad \sum_{x_i} b_i(x_i) = 1 \quad \Big\}$$

# Second attempt

$$b \in \text{LOC}(G)$$

$$\begin{aligned} \text{LOC}(G) &= \Big\{ b = \{b_i, b_{ij}\} : \ \text{locally consistent marginals} \Big\}, \\ & \Big\{ \{b_i, b_{ij}\} : \quad \sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i), \\ & \qquad\qquad\qquad \sum_{x_i} b_i(x_i) = 1 \quad \Big\} \end{aligned}$$

# Geometric picture




Polytopes

# Bethe Free Energy

$$\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$$

Intuition

$$\begin{aligned} \mathbb{F}(b) &\approx \mathbb{E}_b \log \psi_{\rm tot}(x) + H(b) \\ &\approx \text{Energy} + \text{Entropy}\,. \end{aligned}$$

# Bethe Free Energy

$$\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$$

Intuition

$$\begin{aligned} \mathbb{F}(b) &\approx \mathbb{E}_b \log \psi_{\rm tot}(x) + H(b) \\ &\approx \text{Energy} + \text{Entropy}\,. \end{aligned}$$

# Bethe Free Energy

$$\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$$

Intuition

$$\begin{aligned} \mathbb{F}(b) &\approx \mathbb{E}_b \log \psi_{\mathrm{tot}}(x) + H(b) \\ &\approx \text{Energy} + \text{Entropy}\,. \end{aligned}$$

# Energy

$$
\begin{aligned}
\mathbb{E}_b \log \psi_{\rm tot}(x) &= \sum_{(i,j)\in E} \mathbb{E}_b \log \psi_{ij}(x_i, x_j) \\
&= \sum_{(i,j)\in E} \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) \\
&= \sum_{(i,j)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)
\end{aligned}
$$

# Entropy?

# Entropy?

**Idea:** Consider $G$ a tree.

## Proposition

*If $G$ is a tree, then*

$$\mu(x) = \prod_{(i,j)\in E} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i) \, .$$

## Corollary

*If $G$ is a tree, then*

$$H(\mu) = \sum_{i\in V} H(\mu_i) - \sum_{(i,j)\in E} I(\mu_{ij}) \, .$$

# Entropy?

**Idea:** Consider $G$ a tree.

### Proposition

*If $G$ is a tree, then*

$$\mu(x) = \prod_{(i,j)\in E} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,.$$

### Corollary

*If $G$ is a tree, then*

$$H(\mu) = \sum_{i\in V} H(\mu_i) - \sum_{(i,j)\in E} I(\mu_{ij})\,.$$

# Mutual Information

$$
\begin{aligned}
I(\mu_{1,2}) &= \sum_{x_1,x_2} \mu_{1,2}(x_1,x_2) \log \frac{\mu_{1,2}(x_1,x_2)}{\mu_1(x_1)\mu_2(x_2)} \\
&= H(\mu_1) + H(\mu_2) - H(\mu_{1,2})\,.
\end{aligned}
$$

# Proof of the proposition

By induction over $n$.

$n = 1$: Trivial

# Proof of the proposition

By induction over $n$.

$n = 1$: Trivial

# Proof of the proposition

True for $n$, $V = [n]$, new vertex $i = n+1$, connected to $j = n$

$$\begin{aligned}
\mu(x_V, x_n) &= \mu(x_V)\,\mu(x_{n+1}|x_V) \\
&= \mu(x_V)\,\mu(x_{n+1}|x_n) && \text{[Markov]} \\
&= \mu(x_V)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1}) && \text{[Bayes]} \\
&= \prod_{(i,j)\neq(n,n+1)} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1})
\end{aligned}$$

QED

# Proof of the proposition

True for $n$, $V = [n]$, new vertex $i = n+1$, connected to $j = n$

$$
\begin{aligned}
\mu(x_V, x_n) &= \mu(x_V)\,\mu(x_{n+1}|x_V) \\
&= \mu(x_V)\,\mu(x_{n+1}|x_n) && \text{[Markov]} \\
&= \mu(x_V)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1}) && \text{[Bayes]} \\
&= \prod_{(i,j)\neq(n,n+1)} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1})
\end{aligned}
$$

QED

# Proof of the proposition

True for $n$, $V = [n]$, new vertex $i = n + 1$, connected to $j = n$

$$
\begin{aligned}
\mu(x_V, x_n) &= \mu(x_V)\,\mu(x_{n+1}|x_V) \\
&= \mu(x_V)\,\mu(x_{n+1}|x_n) && \text{[Markov]} \\
&= \mu(x_V)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1}) && \text{[Bayes]} \\
&= \prod_{(i,j)\neq(n,n+1)} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1})
\end{aligned}
$$

QED

# Proof of the proposition

True for $n$, $V = [n]$, new vertex $i = n+1$, connected to $j = n$

$$\begin{aligned}
\mu(x_V, x_n) &= \mu(x_V)\,\mu(x_{n+1}|x_V) \\
&= \mu(x_V)\,\mu(x_{n+1}|x_n) && \text{[Markov]} \\
&= \mu(x_V)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1}) && \text{[Bayes]} \\
&= \prod_{(i,j)\neq(n,n+1)} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1})
\end{aligned}$$

QED

# Proof of the proposition

True for $n$, $V = [n]$, new vertex $i = n+1$, connected to $j = n$

$$\begin{aligned}
\mu(x_V, x_n) &= \mu(x_V)\,\mu(x_{n+1}|x_V) \\
&= \mu(x_V)\,\mu(x_{n+1}|x_n) && \text{[Markov]} \\
&= \mu(x_V)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1}) && \text{[Bayes]} \\
&= \prod_{(i,j)\neq(n,n+1)} \frac{\mu_{i,j}(x_i, x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in V} \mu_i(x_i)\,\frac{\mu(x_n, x_{n+1})}{\mu(x_n)\mu(x_{n+1})}\mu(x_{n+1})
\end{aligned}$$

QED

# Let's just export it

$$H_{\text{Bethe}} : \text{LOC}(G) \to \mathbb{R}\,.$$

$$H_{\text{Bethe}}(b) = \sum_{i \in V} H(b_i) - \sum_{(i,j) \in E} I(b_{ij})\,.$$

# Putting everything together

$$
\begin{aligned}
\mathbb{F}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H_{\rm Bethe}(b) \\
&= \sum_{(i,j)\in E} \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i,x_j) - \sum_{(i,j)\in E} I(b_{ij}) + \sum_{i\in V} H(b_i) \\
&= \sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j) \log \psi_{ij}(x_i,x_j) \\
&\quad - \sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j) \log \frac{b_{ij}(x_i,x_j)}{b_i(x_i)b_j(x_j)} - \sum_{i\in V} \sum_{x_i} b_i(x_i) \log b_i(x_i)
\end{aligned}
$$

# Putting everything together

$$
\begin{aligned}
\mathbb{F}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H_{\rm Bethe}(b) \\
&= \sum_{(i,j)\in E} \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) - \sum_{(i,j)\in E} I(b_{ij}) + \sum_{i\in V} H(b_i) \\
&= \sum_{(i,j)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j) \\
&\quad - \sum_{(i,j)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \frac{b_{ij}(x_i, x_j)}{b_i(x_i) b_j(x_j)} - \sum_{i\in V} \sum_{x_i} b_i(x_i) \log b_i(x_i)
\end{aligned}
$$

# Putting everything together

$$\begin{aligned}
\mathbb{F}(b) &= \mathbb{E}_b \log \psi_{\rm tot}(x) + H_{\rm Bethe}(b) \\
&= \sum_{(i,j)\in E} \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) - \sum_{(i,j)\in E} I(b_{ij}) + \sum_{i\in V} H(b_i) \\
&= \sum_{(i,j)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j) \\
&\quad - \sum_{(i,j)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \frac{b_{ij}(x_i, x_j)}{b_i(x_i) b_j(x_j)} - \sum_{i\in V} \sum_{x_i} b_i(x_i) \log b_i(x_i)
\end{aligned}$$

# Putting everything together

$$
\begin{aligned}
\mathbb{F}(b) \;=\;& \sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j) \log \psi_{ij}(x_i,x_j) \\
& - \sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j) \log \; b_{ij}(x_i,x_j) \\
& - \sum_{i\in V} (1-\deg(i)) \sum_{x_i} b_i(x_i) \log b_i(x_i)
\end{aligned}
$$

# Problem

Want to maximize $\mathbb{F}(b)$.

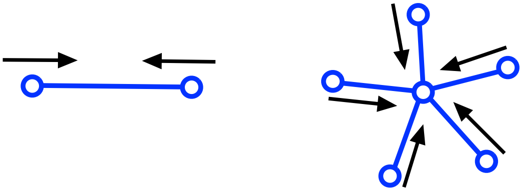
$\mathbb{F}(b)$ is not concave.

# Remark

Assume $\nu$ a BP fixed point

$$\nu_{i\to j}(x_i) \cong \prod_{k\in\partial i\setminus j} \Big\{ \sum_{x_k\in\mathcal{X}} \psi_{ik}(x_i,x_k)\, \nu_{k\to i}(x_k) \Big\}\,.$$

# Define (as you would do on a tree)



$$b_i(x_i) \cong \prod_{k\in\partial i\setminus j} \Big\{ \sum_{x_k\in\mathcal{X}} \psi_{ik}(x_i,x_k)\,\nu_{k\to i}(x_k)\Big\}\,,$$

$$b_{ij}(x_i,x_j) \cong \nu_{i\to j}(x_i)\,\psi_{ij}(x_i,x_j)\,\nu_{j\to i}(x_j)\,.$$

**Lemma**

*With these definitions,* $b \in \mathrm{LOC}(G)$.

# Stationarity condition

Lagrangian

$$\begin{aligned}\mathcal{L}(b,\lambda) &= \mathbb{F}(b) - \sum_{i\in V}\lambda_i\Big\{\sum_{x_i} b_i(x_i) - 1\Big\}\\ &- \sum_{(i,j)\in\vec{E}}\sum_{x_i}\lambda_{i\to j}(x_i)\Big\{\sum_{x_j} b_{ij}(x_i,x_j) - b_i(x_i)\Big\}\end{aligned}$$

$$\nabla_{b_{ij}}\mathcal{L}(b,\lambda) = -1 - b_{ij}(x_i,x_j) + \log\psi_{ij}(x_i,x_j) - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_i)\,,$$

$$\nabla_{b_i}\mathcal{L}(b,\lambda) = -(1-\deg(i))\log[b_i(x_i)\,e] - \lambda_i + \sum_{j\in\partial i}\lambda_{i\to j}(x_i)$$

# Stationarity condition

Lagrangian

$$\begin{aligned}
\mathcal{L}(b,\lambda) &= \mathbb{F}(b) - \sum_{i\in V} \lambda_i \Big\{ \sum_{x_i} b_i(x_i) - 1 \Big\} \\
&\quad - \sum_{(i,j)\in \vec{E}} \sum_{x_i} \lambda_{i\to j}(x_i) \Big\{ \sum_{x_j} b_{ij}(x_i,x_j) - b_i(x_i) \Big\}
\end{aligned}$$

$$\nabla_{b_{ij}} \mathcal{L}(b,\lambda) = -1 - b_{ij}(x_i,x_j) + \log \psi_{ij}(x_i,x_j) - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_i)\,,$$

$$\nabla_{b_i} \mathcal{L}(b,\lambda) = -(1-\deg(i)) \log[b_i(x_i)\, e] - \lambda_i + \sum_{j\in\partial i} \lambda_{i\to j}(x_i)$$

# Stationarity condition

Lagrangian

$$\begin{aligned}\mathcal{L}(b,\lambda) &= \mathbb{F}(b) - \sum_{i\in V}\lambda_i\Big\{\sum_{x_i} b_i(x_i) - 1\Big\} \\ &\quad - \sum_{(i,j)\in\vec{E}}\sum_{x_i}\lambda_{i\to j}(x_i)\Big\{\sum_{x_j} b_{ij}(x_i,x_j) - b_i(x_i)\Big\}\end{aligned}$$

$$\nabla_{b_{ij}}\mathcal{L}(b,\lambda) = -1 - b_{ij}(x_i,x_j) + \log\psi_{ij}(x_i,x_j) - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_i)\,,$$

$$\nabla_{b_i}\mathcal{L}(b,\lambda) = -(1-\deg(i))\log[b_i(x_i)\,e] - \lambda_i + \sum_{j\in\partial i}\lambda_{i\to j}(x_i)$$

# Stationarity condition

$$b_{ij}(x_i, x_j) \;=\; \psi_{ij}(x_i, x_j)\, \exp\big\{-1 - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_j)\big\}\,,$$

$$b_i(x_i) \;\cong\; \exp\Big\{-\frac{1}{\deg(i)-1}\sum_{j\in\partial i}\lambda_{i\to j}(x_i)\Big\}$$

$$\sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i)$$

Did you recognize this?

# Stationarity condition

$$b_{ij}(x_i, x_j) = \psi_{ij}(x_i, x_j) \exp\{ -1 - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_j)\},$$

$$b_i(x_i) \cong \exp\Big\{ -\frac{1}{\deg(i)-1} \sum_{j\in\partial i} \lambda_{i\to j}(x_i)\Big\}$$

$$\sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i)$$

Did you recognize this?

# Defining

$$\nu_{i\to j}(x_i) \;\; \cong \;\; e^{-\lambda_{i\to j}(x_i)}\,.$$

We get

$$b_i(x_i) \;\; \cong \;\; \prod_{k\in\partial i\setminus j}\Big\{\sum_{x_k\in\mathcal{X}}\psi_{ik}(x_i,x_k)\,\nu_{k\to i}(x_k)\Big\}\,,$$

$$b_{ij}(x_i,x_j) \;\; \cong \;\; \nu_{i\to j}(x_i)\,\psi_{ij}(x_i,x_j)\,\nu_{j\to i}(x_j)\,,$$

$$+\text{Local Consistency}$$

# Defining

$$\nu_{i\to j}(x_i) \cong e^{-\lambda_{i\to j}(x_i)}\,.$$

We get

$$\begin{aligned} b_i(x_i) &\cong \prod_{k\in\partial i\setminus j} \Big\{ \sum_{x_k\in\mathcal{X}} \psi_{ik}(x_i,x_k)\,\nu_{k\to i}(x_k)\Big\}\,,\\ b_{ij}(x_i,x_j) &\cong \nu_{i\to j}(x_i)\,\psi_{ij}(x_i,x_j)\,\nu_{j\to i}(x_j)\,,\\ &\quad +\text{Local Consistency} \end{aligned}$$

# We proved the following

**Theorem (Yedidia, Freeman, Weiss, 2003)**

*Fixed points of BP are in one-to-one correspondence with stationary points of Bethe free energy.*

*Fixed point messages are (exponentials of) the dual parameters at the fixed point.*

# We proved the following

**Theorem (Yedidia, Freeman, Weiss, 2003)**

*Fixed points of BP are in one-to-one correspondence with stationary points of Bethe free energy.*

*Fixed point messages are (exponentials of) the dual parameters at the fixed point.*

# Uses

- Alternative algorithms to find fixed points (e.g. gradient ascent).
  [e.g. Heskes 2002]
- Include higher order marginals.
  [Yedidia, Freeman, Weiss, 2003]
- Convexify Bethe free energy.
  [Wainwright, Jaakkola, Willsky, 2005]
- Asymptotically tight estimates on $\log Z$ for graph sequences.
  [e.g. Dembo, Montanari 2010]

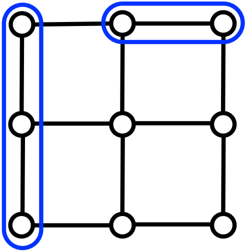
# Region-Based Approximation

# Idea

| | | | | |
|---|---|---|---|---|
| Vertices | $\rightarrow$ | Edges | $\rightarrow$ | Regions |
| Naive Mean Field | $\rightarrow$ | Bethe Free Energy | $\rightarrow$ | Region-Based Free Energy |
| MF Equations | $\rightarrow$ | Belief Propagation | $\rightarrow$ | Generalized BP |

[Cluster variational method, Kikuchi 1951]

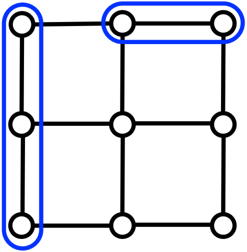
# Region



$R = (V_R, E_R)$, s.t.

- If $(i, j) \in E_R$ then $i, j \in V_R$

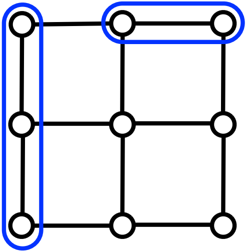
# Region



Free energy of region $R$: $\mathbb{F}_R : \mathsf{M}(\mathcal{X}^{V_R}) \to \mathbb{R}$

$$
\begin{aligned}
\mathbb{F}_R(b_R) &= \mathbb{E}_{b_R} \log \psi_{\mathtt{tot},R}(x_R) + H(b_R) \\
&= \sum_{x_R} \sum_{(i,j)\in E_R} b_R(x_R) \log \psi_{ij}(x_i, x_j) - \sum_{x_R} b_R(x_R) \log b_R(x_R) \, .
\end{aligned}
$$

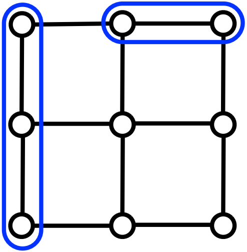
# Region



Free energy of region $R$: $\mathbb{F}_R : \mathsf{M}(\mathcal{X}^{V_R}) \to \mathbb{R}$

$$\begin{aligned} \mathbb{F}_R(b_R) &= \mathbb{E}_{b_R} \log \psi_{\texttt{tot},R}(x_R) + H(b_R) \\ &= \sum_{x_R} \sum_{(i,j)\in E_R} b_R(x_R) \log \psi_{ij}(x_i, x_j) - \sum_{x_R} b_R(x_R) \log b_R(x_R) \,. \end{aligned}$$

# Region



Free energy of region $R$: $\mathbb{F}_R : \mathsf{M}(\mathcal{X}^{V_R}) \to \mathbb{R}$

Can be evaluated for small regions (complexity $|\mathcal{X}|^{|R|}$).

# Region-based Approximation

**Collection of regions**

$$\mathbf{R} = \{R_1, R_2, \dots, R_m\}\,.$$

**Coefficients**

$$c_{\mathbf{R}} = \{c_{R_1}, c_{R_2}, \dots, c_{R_m}\}\,, \qquad c_{R_i} \in \mathbb{R}\,.$$

**Free Energy approximation:**

$$\mathbb{F}_{\mathbf{R}} : \qquad \mathsf{M}(\mathcal{X}^{V(R_1)}) \times \cdots \times \mathsf{M}(\mathcal{X}^{V(R_m)}) \to \mathbb{R}$$

$$b_{\mathbf{R}} = (b_{R_1}, \dots, b_{R_m}) \mapsto \mathbb{F}_{\mathbf{R}}(b_{\mathbf{R}}) = \sum_{R \in \mathbf{R}} c_R\, \mathbb{F}_R(b_R)$$

# Region-based Approximation

**Collection of regions**

$$\mathbf{R} = \{R_1, R_2, \dots, R_m\}\,.$$

**Coefficients**

$$c_{\mathbf{R}} = \{c_{R_1}, c_{R_2}, \dots, c_{R_m}\}\,, \qquad c_{R_i} \in \mathbb{R}\,.$$

**Free Energy approximation:**

$$\mathbb{F}_{\mathbf{R}}: \qquad \mathsf{M}(\mathcal{X}^{V(R_1)}) \times \cdots \times \mathsf{M}(\mathcal{X}^{V(R_m)}) \to \mathbb{R}$$

$$b_{\mathbf{R}} = (b_{R_1}, \dots, b_{R_m}) \mapsto \mathbb{F}_{\mathbf{R}}(b_{\mathbf{R}}) = \sum_{R \in \mathbf{R}} c_R\, \mathbb{F}_R(b_R)$$

# Region-based Approximation

**Collection of regions**

$$\mathbf{R} = \{R_1, R_2, \dots, R_m\}\,.$$

**Coefficients**

$$c_{\mathbf{R}} = \{c_{R_1}, c_{R_2}, \dots, c_{R_m}\}\,, \qquad c_{R_i} \in \mathbb{R}\,.$$

**Free Energy approximation:**

$$\mathbb{F}_{\mathbf{R}}: \qquad \mathsf{M}(\mathcal{X}^{V(R_1)}) \times \cdots \times \mathsf{M}(\mathcal{X}^{V(R_m)}) \to \mathbb{R}$$

$$b_{\mathbf{R}} = (b_{R_1}, \dots, b_{R_m}) \mapsto \mathbb{F}_{\mathbf{R}}(b_{\mathbf{R}}) = \sum_{R \in \mathbf{R}} c_R \, \mathbb{F}_R(b_R)$$

# Region-based Approximation

**Collection of regions**

$$\mathbf{R} = \{R_1, R_2, \dots, R_m\}\,.$$

**Coefficients**

$$c_{\mathbf{R}} = \{c_{R_1}, c_{R_2}, \dots, c_{R_m}\}\,, \qquad c_{R_i} \in \mathbb{R}\,.$$

**Free Energy approximation:**

$$\mathbb{F}_{\mathbf{R}}: \qquad \mathsf{M}(\mathcal{X}^{V(R_1)}) \times \cdots \times \mathsf{M}(\mathcal{X}^{V(R_m)}) \to \mathbb{R}$$

$$b_{\mathbf{R}} = (b_{R_1}, \dots, b_{R_m}) \mapsto \mathbb{F}_{\mathbf{R}}(b_{\mathbf{R}}) = \sum_{R \in \mathbf{R}} c_R\, \mathbb{F}_R(b_R)$$

# Example: Bethe Free Energy

**Regions**

$$\begin{aligned} \mathbf{R} &= \{R_i : i \in V\} \cup \{R_{ij} : (i,j) \in E\}, \\ R_i &= (\{i\}, \emptyset), \\ R_{ij} &= (\{i,j\}, \{(i,j)\}). \end{aligned}$$

**Coefficients**

$$c_i = 1 - \deg(i), \qquad c_{ij} = 1.$$

**Free energy**

$$\mathbb{F}_{\mathbf{R}}(b) = \sum_{i \in V} \{1 - \deg(i)\} H(b_i) + \sum_{(i,j) \in E} \Big\{ H(b_{ij}) + \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) \Big\}$$

# Example: Bethe Free Energy

**Regions**

$$
\begin{aligned}
\mathbf{R} &= \{R_i : i \in V\} \cup \{R_{ij} : (i,j) \in E\}\,, \\
R_i &= (\{i\}, \emptyset)\,, \\
R_{ij} &= (\{i,j\}, \{(i,j)\})\,.
\end{aligned}
$$

**Coefficients**

$$c_i = 1 - \deg(i)\,, \qquad c_{ij} = 1\,.$$

**Free energy**

$$\mathbb{F}_{\mathbf{R}}(b) = \sum_{i \in V} \{1 - \deg(i)\}\, H(b_i) + \sum_{(i,j)\in E} \Big\{ H(b_{ij}) + \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) \Big\}$$

# Example: Bethe Free Energy

**Regions**

$$\begin{aligned}\mathbf{R} &= \{R_i:\ i\in V\}\cup\{R_{ij}:\ (i,j)\in E\}\,,\\ R_i &= (\{i\},\emptyset)\,,\\ R_{ij} &= (\{i,j\},\{(i,j)\})\,.\end{aligned}$$

**Coefficients**

$$c_i = 1-\deg(i)\,,\qquad c_{ij}=1\,.$$

**Free energy**

$$\mathbb{F}_{\mathbf{R}}(b) = \sum_{i\in V}\{1-\deg(i)\}\,H(b_i) + \sum_{(i,j)\in E}\Big\{H(b_{ij}) + \mathbb{E}_{b_{ij}}\log\psi_{ij}(x_i,x_j)\Big\}$$

# Example: Bethe Free Energy

**Regions**

$$\begin{aligned} \mathbf{R} &= \{R_i : \ i \in V\} \cup \{R_{ij} : \ (i,j) \in E\}\,, \\ R_i &= (\{i\}, \emptyset)\,, \\ R_{ij} &= (\{i,j\}, \{(i,j)\})\,. \end{aligned}$$

**Coefficients**

$$c_i = 1 - \deg(i)\,, \qquad c_{ij} = 1\,.$$

**Free energy**

$$\mathbb{F}_{\mathbf{R}}(b) = \sum_{i \in V} \{1 - \deg(i)\}\, H(b_i) + \sum_{(i,j)\in E} \Big\{ H(b_{ij}) + \mathbb{E}_{b_{ij}} \log \psi_{ij}(x_i, x_j) \Big\}$$

# Questions?

1. What about domain/consistency?

2. How to choose coefficients?

3. How to choose regions?

# Valid Region-Based Approximations

[Yedidia, Freeman, Weiss, 2003]

**Condition 1:** Consistency

$$R \in \mathbf{R},\ R' \subseteq R \quad \Rightarrow \quad R' \in \mathbf{R}\,.$$

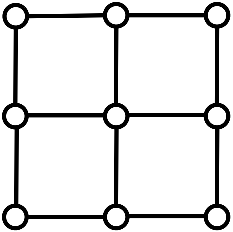
**Condition 2:** Vertex counting

$$\sum_{R \in \mathbf{R}} c_R\, \mathbb{I}(i \in R) = 1 \qquad \text{for all } i \in V\,.$$

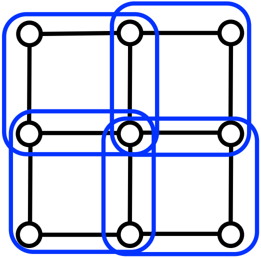
**Condition 3:** Edge counting

$$\sum_{R \in \mathbf{R}} c_R\, \mathbb{I}((i,j) \in R) = 1 \qquad \text{for all } (i,j) \in E\,.$$

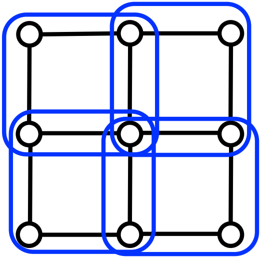
# Example



# Example

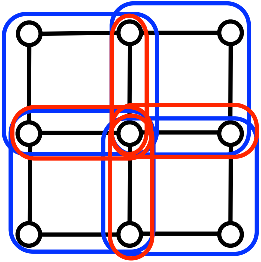


# Example



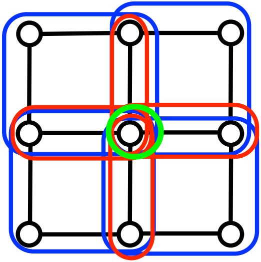
Add intersections!

# Example



Add intersections!

# Example



Add intersections!

# Why? #1

$$R \in \mathbf{R},\, R' \subseteq R \quad \Rightarrow \quad R' \in \mathbf{R}\,.$$

Clean local consistecy conditions

$$\sum_{x_{R\setminus R'}} b_R(x_R) = b_{R'}(x_{R'}) \qquad \text{for all } R' \subseteq R\,.$$

$\mathrm{LOC}(G;\mathbf{R})$

# Why? #1

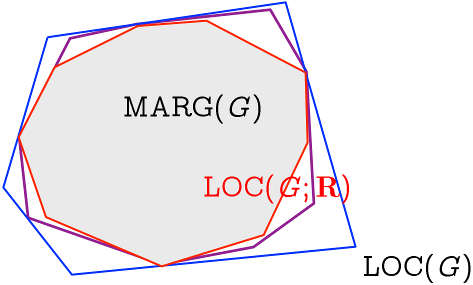
$$R \in \mathbf{R},\ R' \subseteq R \quad \Rightarrow \quad R' \in \mathbf{R}\,.$$

Clean local consistecy conditions

$$\sum_{x_{R\setminus R'}} b_R(x_R) = b_{R'}(x_{R'}) \qquad \text{for all } R' \subseteq R\,.$$

$\mathrm{LOC}(G;\mathbf{R})$

# Geometric picture




Polytopes

# Why? #2

$$\sum_{R\in\mathbf{R}} c_R\,\mathbb{I}(i\in R) = 1 \qquad \text{for all } i\in V\,.$$

Consider $\psi_{ij}(x_i,x_j)=1$, $b_R(x_R)=$ Uniform

$$\begin{aligned}\sum_{R\in\mathbf{R}} \mathbb{F}_R(b_R) &= \sum_{R\in\mathbf{R}} c_R\,H(b_R)\\ &= \sum_{R\in\mathbf{R}} c_R\,|V(R)|\log|\mathcal{X}|\\ &= \sum_{i\in V}\Big\{\sum_{R\in\mathbf{R}} c_R\,\mathbb{I}(i\in R)\Big\}\log|\mathcal{X}| = |V|\log|\mathcal{X}|\end{aligned}$$

# Why? #2

$$\sum_{R \in \mathbf{R}} c_R \, \mathbb{I}(i \in R) = 1 \qquad \text{for all } i \in V\,.$$

Consider $\psi_{ij}(x_i, x_j) = 1$, $b_R(x_R) = \text{Uniform}$

$$\begin{aligned}
\sum_{R \in \mathbf{R}} \mathbb{F}_R(b_R) &= \sum_{R \in \mathbf{R}} c_R \, H(b_R) \\
&= \sum_{R \in \mathbf{R}} c_R \, |V(R)| \log |\mathcal{X}| \\
&= \sum_{i \in V} \left\{ \sum_{R \in \mathbf{R}} c_R \, \mathbb{I}(i \in R) \right\} \log |\mathcal{X}| = |V| \log |\mathcal{X}|
\end{aligned}$$

# Why? #3

$$\sum_{R\in\mathbf{R}} c_R\,\mathbb{I}((i,j)\in R) = 1 \qquad \text{for all } (i,j)\in E\,.$$

Neglect entropy (e.g. $\psi_{ij}(x_i,x_j) = e^{\beta\,\theta_{ij}(x_i,x_j)}$, $\beta\to\infty$)

$$\begin{aligned}
\sum_{R\in\mathbf{R}} \mathbb{F}_R(b_R) &= \beta\sum_{R\in\mathbf{R}} c_R \sum_{x_R} b_R(x_R) \sum_{(ij)\in E(R)} \theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta\sum_{R\in\mathbf{R}} c_R \sum_{(ij)\in E(R)} \mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta\sum_{(ij)\in E}\left\{\sum_{R\in\mathbf{R}} c_R\,\mathbb{I}((i,j)\in R)\right\}\mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta\sum_{(ij)\in E} \mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)
\end{aligned}$$

# Why? #3

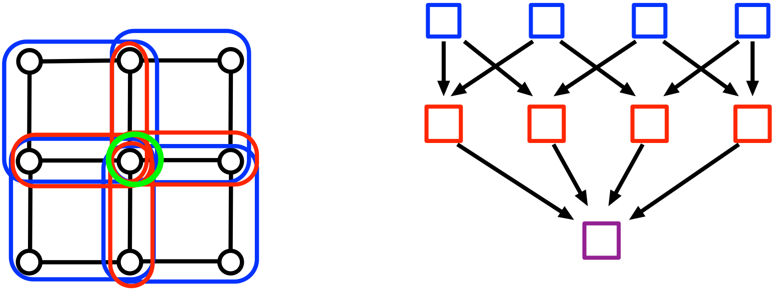
$$\sum_{R\in\mathbf{R}} c_R\,\mathbb{I}((i,j)\in R) = 1 \qquad \text{for all } (i,j)\in E\,.$$

Neglect entropy (e.g. $\psi_{ij}(x_i,x_j) = e^{\beta\,\theta_{ij}(x_i,x_j)}$, $\beta\to\infty$)

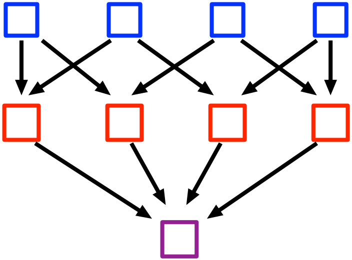
$$\begin{aligned}
\sum_{R\in\mathbf{R}} \mathbb{F}_R(b_R) &= \beta \sum_{R\in\mathbf{R}} c_R \sum_{x_R} b_R(x_R) \sum_{(ij)\in E(R)} \theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta \sum_{R\in\mathbf{R}} c_R \sum_{(ij)\in E(R)} \mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta \sum_{(ij)\in E} \left\{ \sum_{R\in\mathbf{R}} c_R\,\mathbb{I}((i,j)\in R) \right\} \mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)\\
&= \beta \sum_{(ij)\in E} \mathbb{E}_{b_{ij}}\theta_{ij}(x_i,x_j) + O_\beta(1)
\end{aligned}$$

## How do you compute the coefficients?

# The Region Graph



# The Region Graph



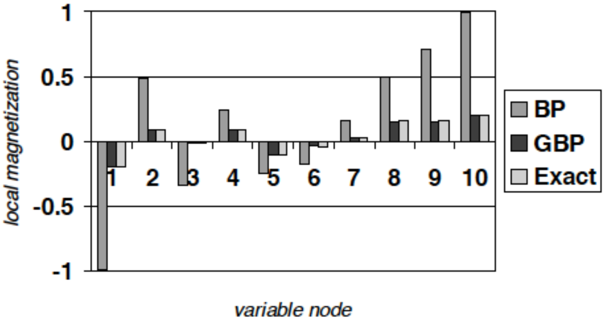
$c_{\text{loop}} = +1$

$c_{\text{edge}} = +1$

$c_{\text{vert}} = +1$

$$c_R = 1 - \sum_{R' \in \text{ANCESTORS}(R)} c_R$$

# Was it worth it?




$10 \times 10$ Ising model with random potentials [Yedidia et al. 2003]

# Tree-Based Convexifications

# Intermezzo: Exponential Families

$$T : \quad \mathcal{X}^V \to \mathbb{R}^m ,$$
$$x \mapsto T(x) = (T_1(x), \dots, T_m(x)) .$$

Exponential family $\{\mu_\theta : \theta \in \mathbb{R}^m\}$

$$\mu_\theta(x) = \frac{1}{Z(\theta)} \exp\Big\{\langle \theta, T(x)\rangle\Big\} , \qquad F(\theta) = \log Z(\theta)$$

# Intermezzo: Exponential Families

$$T \; : \quad \mathcal{X}^V \to \mathbb{R}^m ,$$
$$x \mapsto T(x) = (T_1(x), \dots, T_m(x)) .$$

Exponential family $\{\mu_\theta \; : \; \theta \in \mathbb{R}^m\}$

$$\mu_\theta(x) = \frac{1}{Z(\theta)} \exp\Big\{\langle \theta, T(x)\rangle\Big\} , \qquad F(\theta) = \log Z(\theta)$$

# Exponential Families: Basic Properties

## Proposition

$$
\begin{aligned}
&(1) && \theta \mapsto F(\theta) \text{ is convex,} \\
&(2) && \nabla_\theta F(\theta) = \mathbb{E}_\theta\{T(x)\} \equiv \tau(\theta)\,, \\
&(3) && \nabla^2_\theta F(\theta) = \mathrm{Cov}_\theta\{T(x); T(x))\,, \\
&(4) && \overline{\mathrm{Image}(\tau)} = \mathrm{MARG}(T)\,.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{MARG}(T) &\equiv \mathrm{conv}\Big(\{T(x) : x \in \mathcal{X}^V\}\Big) \\
&= \Big\{\mathbb{E}_\nu T(x) : \nu \in \mathrm{M}(\mathcal{X}^V)\Big\}
\end{aligned}
$$

# Exponential Families: Basic Properties

## Proposition

$$\begin{aligned}
&(1) && \theta \mapsto F(\theta) \textit{ is convex}, \\
&(2) && \nabla_\theta F(\theta) = \mathbb{E}_\theta\{T(x)\} \equiv \tau(\theta)\,, \\
&(3) && \nabla^2_\theta F(\theta) = \mathrm{Cov}_\theta\{T(x);\, T(x)\}\,, \\
&(4) && \overline{\mathrm{Image}(\tau)} = \mathrm{MARG}(T)\,.
\end{aligned}$$

$$\mathrm{MARG}(T) \equiv \mathrm{conv}\Big(\{T(x):\ x \in \mathcal{X}^V\}\Big)$$

$$= \Big\{\mathbb{E}_\nu T(x):\ \nu \in \mathsf{M}(\mathcal{X}^V)\Big\}$$

# Exponential Families: Basic Properties

## Proposition

$$\begin{aligned}
&(1) \qquad \theta \mapsto F(\theta) \ \textit{is convex}, \\
&(2) \qquad \nabla_\theta F(\theta) = \mathbb{E}_\theta\{T(x)\} \equiv \tau(\theta)\,, \\
&(3) \qquad \nabla^2_\theta F(\theta) = \mathrm{Cov}_\theta\{T(x); T(x))\,, \\
&(4) \qquad \overline{\mathrm{Image}(\tau)} = \mathrm{MARG}(T)\,.
\end{aligned}$$

$$\begin{aligned}
\mathrm{MARG}(T) &\equiv \mathrm{conv}\Big(\{T(x): \ x \in \mathcal{X}^V\}\Big) \\
&= \Big\{\mathbb{E}_\nu T(x): \ \nu \in \mathsf{M}(\mathcal{X}^V)\Big\}
\end{aligned}$$

# Proofs

**(1), (2), (3):** Exercises

(4): A bit more difficult

**Claim 1:** A closed convex set is the closure of its relative interior.
[Hint: Assume the set has full dimension. Each point has a cone of full dimension around it.]

**Claim 2:** Let $\tau_* \in \text{relint}(\text{MARG}(T))$. Then $\tau_* = \mathbb{E}_{\nu_*}\{T(x)\}$ for some $\nu_*$ s.t. $\nu_*(x) > 0$ for all $x \in \mathcal{X}^V$.
[Hint: Consider the set of signed weigths $\nu$ such that $\sum_x \nu(x) T(x) = \tau_*$. If the claim was false, it would be tangent to the simplex.]

**Claim 3:** There exists $\theta_* \in \mathbb{R}^m$ such that $\mathbb{E}_{\theta_*}\{T(x)\} = \mathbb{E}_{\nu_*}\{T(x)\}$.

# Proofs

**(1), (2), (3):** Exercises

**(4):** A bit more difficult

**Claim 1:** A closed convex set is the closure of its relative interior.
[Hint: Assume the set has full dimension. Each point has a cone of full dimension around it.]

**Claim 2:** Let $\tau_* \in \text{relint}(\text{MARG}(T))$. Then $\tau_* = \mathbb{E}_{\nu_*}\{T(x)\}$ for some $\nu_*$ s.t. $\nu_*(x) > 0$ for all $x \in \mathcal{X}^V$.
[Hint: Consider the set of signed weigths $\nu$ such that $\sum_x \nu(x)T(x) = \tau_*$. If the claim was false, it would be tangent to the simplex.]

**Claim 3:** There exists $\theta_* \in \mathbb{R}^m$ such that $\mathbb{E}_{\theta_*}\{T(x)\} = \mathbb{E}_{\nu_*}\{T(x)\}$.

# Proof of Claim 3

Wlog $\{1, T_1, \dots, T_m\}$ linearly independent.
Consider

$$\begin{aligned} F(\theta;\tau_*) &\equiv F(\theta) - \langle \tau_*, \theta\rangle \\ &= \log\Big\{ \sum_{x\in\mathcal{X}^V} \exp\left(\langle \theta, T(x)\rangle\right)\Big\} - \mathbb{E}_{\nu_*}\{\langle \theta, T(x)\rangle\} \end{aligned}$$

- $F(\,\cdot\,;\tau_*) : \mathbb{R}^m \to \mathbb{R}$ is differentiable and convex.
- If $\theta_*$ is a stationary point, then $\mathbb{E}_{\theta_*}\{T(x)\} = \mathbb{E}_{\nu_*}\{T(x)\}$.
- As $\theta \to \infty$, $F(\theta;\tau_*) \to \infty$.

Implies the thesis.

# Proof of Claim 3

Wlog $\{1, T_1, \dots, T_m\}$ linearly independent.
Consider

$$
\begin{aligned}
F(\theta;\tau_*) &\equiv F(\theta) - \langle \tau_*, \theta\rangle \\
&= \log\Big\{\sum_{x\in\mathcal{X}^V} \exp\left(\langle\theta, T(x)\rangle\right)\Big\} - \mathbb{E}_{\nu_*}\{\langle\theta, T(x)\rangle\}
\end{aligned}
$$

- $F(\,\cdot\,;\tau_*) : \mathbb{R}^m \to \mathbb{R}$ is differentiable and convex.
- If $\theta_*$ is a stationary point, then $\mathbb{E}_{\theta_*}\{T(x)\} = \mathbb{E}_{\nu_*}\{T(x)\}$.
- As $\theta \to \infty$, $F(\theta;\tau_*) \to \infty$.

Implies the thesis.

# Proof of Claim 3

Wlog $\{1, T_1, \dots, T_m\}$ linearly independent.
Consider

$$\begin{aligned} F(\theta; \tau_*) &\equiv F(\theta) - \langle \tau_*, \theta \rangle \\ &= \log \Big\{ \sum_{x \in \mathcal{X}^V} \exp\big(\langle \theta, T(x) \rangle\big) \Big\} - \mathbb{E}_{\nu_*}\{\langle \theta, T(x) \rangle\} \end{aligned}$$

- $F(\,\cdot\,; \tau_*) : \mathbb{R}^m \to \mathbb{R}$ is differentiable and convex.
- If $\theta_*$ is a stationary point, then $\mathbb{E}_{\theta_*}\{T(x)\} = \mathbb{E}_{\nu_*}\{T(x)\}$.
- As $\theta \to \infty$, $F(\theta; \tau_*) \to \infty$.

Implies the thesis.

## As $\theta \to \infty$, $F_{\tau_*}(\theta) \to \infty$

Let $\theta = \beta\, v$, $\beta \in \mathbb{R}_+$

$$
\begin{aligned}
F(\theta; \tau_*) &= \log\Big\{ \sum_{x \in \mathcal{X}^V} \exp\left(\langle \theta, T(x) \rangle\right) \Big\} - \mathbb{E}_{\nu_*}\{\langle \theta, T(x) \rangle\} \\
&\geq \beta \Big[ \max_x \langle v, T(x) \rangle - \mathbb{E}_{\nu_*}\{\langle v, T(x) \rangle\} \Big]
\end{aligned}
$$

and $[\ \dots\ ] > 0$ strictly because $\nu_*(x) > 0$ for all $x$.

# Duality structure

$$F_*(\tau) \equiv \inf_{\theta \in \mathbb{R}^m} \{F(\theta) - \langle \tau, \theta \rangle\},$$

$$F_* : \quad \text{MARG}(T) \to \mathbb{R}, \qquad \text{concave.}$$

$$F(\theta) \equiv \sup_{\tau \in \text{MARG}(T)} \{F_*(\tau) + \langle \tau, \theta \rangle\},$$

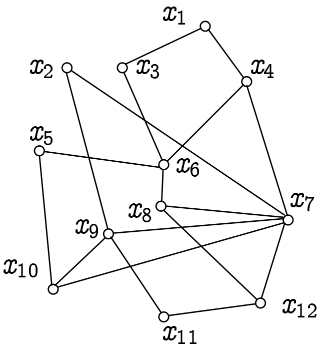
$$F : \quad \mathbb{R}^m \to \mathbb{R}, \qquad \text{convex.}$$

# Duality structure

$$
\begin{aligned}
F_*(\tau) &\equiv \inf_{\theta\in\mathbb{R}^m} \left\{F(\theta) - \langle \tau, \theta\rangle\right\}, \\
F_* : &\quad \mathrm{MARG}(T) \to \mathbb{R}, \qquad \text{concave.}
\end{aligned}
$$

$$
\begin{aligned}
F(\theta) &\equiv \sup_{\tau\in\mathrm{MARG}(T)} \left\{F_*(\tau) + \langle \tau, \theta\rangle\right\}, \\
F : &\quad \mathbb{R}^m \to \mathbb{R}, \qquad \text{convex.}
\end{aligned}
$$

# Let's apply all this




$G = (V, E)$, $V = [n]$, $x = (x_1, \dots, x_n)$, $x_i \in \mathcal{X}$,

$$\begin{aligned} T_{i,\xi}(x) &= \mathbb{I}(x_i = \xi)\,, & i \in V, \xi \in \mathcal{X}, \\ T_{ij,\xi_1,\xi_2}(x) &= \mathbb{I}(x_i = \xi_1)\,\mathbb{I}(x_j = \xi_2)\,, & (i,j) \in E, \xi_1, \xi_2 \in \mathcal{X}, \end{aligned}$$

overcomplete!

# The exponential family

$$\mu_\theta(x) = \frac{1}{Z(\theta)} \exp\left\{ \sum_{(i,j)\in E, \xi_1,\xi_2\in\mathcal{X}} \theta_{ij}(\xi_1,\xi_2)\, T_{ij\xi_1\xi_2}(x) + \sum_{i\in V,\xi\in\mathcal{X}} \theta_i(\xi) T_{i\xi}(x) \right\}$$

$$= \frac{1}{Z(\theta)} \exp\left\{ \sum_{(i,j)\in E} \theta_{ij}(x_i,x_j) + \sum_{i\in V} \theta_i(x_i) \right\}$$

(General pairwise model)

The $\tau$ parameters

$$b_i(\xi) = \mathbb{E}_\theta\{T_i(\xi)\} = \mu_\theta(x_i = \xi), \qquad \text{for } i \in V,$$

$$b_{ij}(\xi_1,\xi_2) = \mathbb{E}_\theta\{T_{ij}(\xi_1,\xi_2)\} = \mu_\theta(x_i = \xi_1, x_j = \xi_2), \qquad \text{for } (i,j) \in E.$$

# The exponential family

$$\mu_\theta(x) = \frac{1}{Z(\theta)} \exp\left\{ \sum_{(i,j)\in E, \xi_1,\xi_2\in\mathcal{X}} \theta_{ij}(\xi_1,\xi_2)\, T_{ij\xi_1\xi_2}(x) + \sum_{i\in V,\xi\in\mathcal{X}} \theta_i(\xi) T_{i\xi}(x) \right\}$$

$$= \frac{1}{Z(\theta)} \exp\left\{ \sum_{(i,j)\in E} \theta_{ij}(x_i,x_j) + \sum_{i\in V} \theta_i(x_i) \right\}$$

(General pairwise model)

The $\tau$ parameters

$$\begin{aligned} b_i(\xi) &= \mathbb{E}_\theta\{T_i(\xi)\} = \mu_\theta(x_i=\xi), && \text{for } i \in V, \\ b_{ij}(\xi_1,\xi_2) &= \mathbb{E}_\theta\{T_{ij}(\xi_1,\xi_2)\} = \mu_\theta(x_i=\xi_1, x_j=\xi_2)\,, && \text{for } (i,j)\in E. \end{aligned}$$

# The duality structure

$$F(\theta) \leftrightarrow F_*(b)\,,$$
$$F_*: \qquad \text{MARG}(G) \to \mathbb{R}\,.$$

We want to evaluate at $\Phi = F(\theta_* = \log\psi)$':

$$\begin{aligned}\Phi &= \sup_{b\in\text{MARG}(G)} \Big\{F_*(b) + \langle\theta_*, b\rangle\Big\}\\ &= \text{Entropy} + \text{Energy}\end{aligned}$$

**New interpretation**

Bethe entropy is an approximate expression for $F_*(b)$.

# The duality structure

$$F(\theta) \leftrightarrow F_*(b)\,,$$
$$F_* : \qquad \mathrm{MARG}(G) \to \mathbb{R}\,.$$

We want to evaluate at $\Phi = F(\theta_* = \log\psi)$':

$$\begin{aligned}\Phi &= \sup_{b\in\mathrm{MARG}(G)} \Big\{F_*(b) + \langle\theta_*, b\rangle\Big\}\\ &= \text{Entropy} + \text{Energy}\end{aligned}$$

## New interpretation

Bethe entropy is an approximate expression for $F_*(b)$.

# The duality structure

$$F(\theta) \leftrightarrow F_*(b)\,,$$
$$F_* : \qquad \text{MARG}(G) \to \mathbb{R}\,.$$

We want to evaluate at $\Phi = F(\theta_* = \log \psi)$':

$$\begin{aligned} \Phi &= \sup_{b \in \text{MARG}(G)} \Big\{ F_*(b) + \langle \theta_*, b \rangle \Big\} \\ &= \text{Entropy} + \text{Energy} \end{aligned}$$

## New interpretation

Bethe entropy is an approximate expression for $F_*(b)$.

# Interpretation works fine on trees

**Proposition**

*If $G$ is a tree, then* $\mathrm{MARG}(G) = \mathrm{LOC}(G)$ *and*

$$F_*(b) = \sum_{i \in V} H(b_i) - \sum_{(i,j) \in E} I(b_{ij}) = \mathbb{F}_{\psi=1}(b)$$

*As a consequence,* $\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$ *is concave.*

**Proof:** Exercise.

# Interpretation works fine on trees

**Proposition**

*If $G$ is a tree, then* $\mathrm{MARG}(G) = \mathrm{LOC}(G)$ *and*

$$F_*(b) = \sum_{i \in V} H(b_i) - \sum_{(i,j) \in E} I(b_{ij}) = \mathbb{F}_{\psi=1}(b)$$

*As a consequence,* $\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$ *is concave.*

**Proof:** Exercise.

# Interpretation works fine on trees

## Proposition

*If $G$ is a tree, then* $\mathrm{MARG}(G) = \mathrm{LOC}(G)$ *and*

$$F_*(b) = \sum_{i\in V} H(b_i) - \sum_{(i,j)\in E} I(b_{ij}) = \mathbb{F}_{\psi=1}(b)$$

*As a consequence,* $\mathbb{F} : \mathrm{LOC}(G) \to \mathbb{R}$ *is concave.*

**Proof:** Exercise.

# What about general graphs?

Write $G$ as a convex combination of trees.

**Abuse:** I will use $T$ to denote trees, not functions.

# Convex combinations

$$\mathcal{T}(G) = \{ \text{spanning trees in } G \},$$

$$\rho : \mathcal{T}(G) \to [0,1],$$
$$T \mapsto \rho_T, \qquad \text{weights},$$

$$\sum_{T \in \mathcal{T}(G)} \rho_T = 1,$$
$$\sum_{T \in \mathcal{T}(G)} \rho_T \, \theta^T = \theta.$$

# Convex combinations

$$\mathcal{T}(G) = \{\text{ spanning trees in } G\,\}\,,$$

$$\begin{aligned}\rho:\mathcal{T}(G) &\to [0,1]\,,\\ T &\mapsto \rho_T\,, \qquad \text{weights}\,,\end{aligned}$$

$$\sum_{T\in\mathcal{T}(G)} \rho_T = 1\,,$$

$$\sum_{T\in\mathcal{T}(G)} \rho_T\,\theta^T = \theta\,.$$

# Convex combinations

$$\mathcal{T}(G) \;=\; \{\,\text{spanning trees in } G\,\}\,,$$

$$\begin{aligned}\rho : \mathcal{T}(G) &\to [0,1]\,,\\ T &\mapsto \rho_T\,, \qquad \text{weights}\,,\end{aligned}$$

$$\sum_{T\in\mathcal{T}(G)} \rho_T = 1\,,$$

$$\sum_{T\in\mathcal{T}(G)} \rho_T\,\theta^T = \theta\,.$$

# Convex combinations

$$\Phi = F(\theta) = F\Big(\sum_{T\in\mathcal{T}(G)} \rho_T \theta^T\Big) \le \sum_{T\in\mathcal{T}(G)} \rho_T\, F(\theta^T)$$

- Fix weigths $\rho_T$.
- Minimize over $\theta^T$ (convex!)

Problem: Exponentially many spanning trees.

# Convex combinations

$$
\begin{aligned}
\Phi &= F(\theta) = F\Big(\sum_{T\in\mathcal{T}(G)} \rho_T \theta^T\Big) \\
&\le \sum_{T\in\mathcal{T}(G)} \rho_T\, F(\theta^T)
\end{aligned}
$$

- Fix weigths $\rho_T$.
- Minimize over $\theta^T$ (convex!)

Problem: Exponentially many spanning trees.

# Convex combinations

$$\begin{aligned}\Phi &= F(\theta) = F\Big(\sum_{T\in\mathcal{T}(G)} \rho_T \theta^T\Big) \\ &\le \sum_{T\in\mathcal{T}(G)} \rho_T\, F(\theta^T)\end{aligned}$$

- Fix weigths $\rho_T$.
- Minimize over $\theta^T$ (convex!)

Problem: Exponentially many spanning trees.

# Minimization over $(\theta^T)_{T\in\mathcal{T}(G)}$

$$\begin{aligned}
\text{minimize} \quad & \sum_{T\in\mathcal{T}(G)} \rho_T\, F(\theta^T)\,, \\
\text{subject to} \quad & \sum_{T\in\mathcal{T}(G)} \rho_T \theta^T_{ij}(x_i,x_j) = \theta_{ij}(x_i,x_j)\,, \\
& \sum_{T\in\mathcal{T}(G)} \rho_T \theta^T_{i}(x_i) = \theta_{i}(x_i)\,.
\end{aligned}$$

**Convex Problem**

# Minimization over $(\theta^T)_{T \in \mathcal{T}(G)}$

$$
\begin{aligned}
\text{minimize} \quad & \sum_{T \in \mathcal{T}(G)} \rho_T \, F(\theta^T)\,, \\
\text{subject to} \quad & \sum_{T \in \mathcal{T}(G)} \rho_T \theta_{ij}^T(x_i, x_j) = \theta_{ij}(x_i, x_j)\,, \\
& \sum_{T \in \mathcal{T}(G)} \rho_T \theta_i^T(x_i) = \theta_i(x_i)\,.
\end{aligned}
$$

**Convex Problem**

# Lagrangian

$$\mathcal{L}((\theta^T), b) = \sum_T \rho_T \, F(\theta^T)$$
$$- \sum_{(ij)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \Big\{ \sum_T \rho_T \theta^T_{ij}(x_i, x_j) - \theta_{ij}(x_i, x_j) \Big\}$$
$$- \sum_{i \in V} \sum_{x_i} b_i(x_i) \Big\{ \sum_T \rho_T \theta^T_i(x_i) - \theta_i(x_i) \Big\}$$

$$= \sum_T \rho_T \Big\{ F(\theta^T) - \langle b, \theta^T \rangle \Big\} + \langle b, \theta \rangle$$

**Separable in $\theta^T$**

# Lagrangian

$$\mathcal{L}((\theta^T), b) = \sum_T \rho_T \, F(\theta^T)$$
$$- \sum_{(ij)\in E} \sum_{x_i,x_j} b_{ij}(x_i, x_j) \Big\{ \sum_T \rho_T \theta^T_{ij}(x_i, x_j) - \theta_{ij}(x_i, x_j) \Big\}$$
$$- \sum_{i\in V} \sum_{x_i} b_i(x_i) \Big\{ \sum_T \rho_T \theta^T_i(x_i) - \theta_i(x_i) \Big\}$$

$$= \sum_T \rho_T \Big\{ F(\theta^T) - \langle b, \theta^T \rangle \Big\} + \langle b, \theta \rangle$$

**Separable in $\theta^T$**

# Lagrangian

$$
\begin{aligned}
\mathcal{L}((\theta^T), b) = &\sum_T \rho_T \, F(\theta^T) \\
&- \sum_{(ij)\in E} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \Big\{ \sum_T \rho_T \theta^T_{ij}(x_i, x_j) - \theta_{ij}(x_i, x_j) \Big\} \\
&- \sum_{i\in V} \sum_{x_i} b_i(x_i) \Big\{ \sum_T \rho_T \theta^T_i(x_i) - \theta_i(x_i) \Big\}
\end{aligned}
$$

$$
= \sum_T \rho_T \Big\{ F(\theta^T) - \langle b, \theta^T \rangle \Big\} + \langle b, \theta \rangle
$$

**Separable in $\theta^T$**

# Lagrangian

$$\min_{(\theta^T)} \mathcal{L}((\theta^T), b) = \sum_T \rho_T F_*(b;\theta) + \langle b, \theta \rangle$$

$$= \sum_T \rho_T \Big\{ \sum_{i\in V} H(b_i) - \sum_{(ij)\in E(T)} I(b_{ij}) \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i\in V} H(b_i) \Big\{ \sum_{T:\, i\in V} \rho_T \Big\} - \sum_{(i,j)\in V} I(b_{ij}) \Big\{ \sum_{T:\,(i,j)\in E(T)} \rho_T \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle$$

# Lagrangian

$$\min_{(\theta^T)} \mathcal{L}((\theta^T), b) = \sum_T \rho_T F_*(b;\theta) + \langle b, \theta \rangle$$

$$= \sum_T \rho_T \Big\{ \sum_{i \in V} H(b_i) - \sum_{(ij) \in E(T)} I(b_{ij}) \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i \in V} H(b_i) \Big\{ \sum_{T:\, i \in V} \rho_T \Big\} - \sum_{(i,j) \in V} I(b_{ij}) \Big\{ \sum_{T:\, (i,j) \in E(T)} \rho_T \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i \in V} H(b_i) - \sum_{(i,j) \in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle$$

# Lagrangian

$$\min_{(\theta^T)} \mathcal{L}((\theta^T), b) = \sum_T \rho_T F_*(b;\theta) + \langle b, \theta \rangle$$

$$= \sum_T \rho_T \Big\{ \sum_{i\in V} H(b_i) - \sum_{(ij)\in E(T)} I(b_{ij}) \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i\in V} H(b_i) \Big\{ \sum_{T:\, i\in V} \rho_T \Big\} - \sum_{(i,j)\in V} I(b_{ij}) \Big\{ \sum_{T:\,(i,j)\in E(T)} \rho_T \Big\} + \langle b, \theta \rangle$$

$$= \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle$$

# Lagrangian

$$
\begin{aligned}
\min_{(\theta^T)} \mathcal{L}((\theta^T), b) &= \sum_T \rho_T F_*(b;\theta) + \langle b, \theta \rangle \\
&= \sum_T \rho_T \Big\{ \sum_{i\in V} H(b_i) - \sum_{(ij)\in E(T)} I(b_{ij}) \Big\} + \langle b, \theta \rangle \\
&= \sum_{i\in V} H(b_i) \Big\{ \sum_{T:\, i\in V} \rho_T \Big\} - \sum_{(i,j)\in V} I(b_{ij}) \Big\{ \sum_{T:\,(i,j)\in E(T)} \rho_T \Big\} + \langle b, \theta \rangle \\
&= \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle
\end{aligned}
$$

# Tree-reweighted free energy

$$\mathbb{F}_{\rm TRW}(b) = \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle$$

Compare with Bethe free energy

$$\mathbb{F}(b) \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} I(b_{ij}) + \langle b, \theta \rangle$$

$\rho(i,j) = 0$ Obviously concave upper bound.

$\rho(i,j) = 1$ Bethe free energy.

# Tree-reweighted free energy

$$\mathbb{F}_{\rm TRW}(b) = \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta\rangle$$

Compare with Bethe free energy

$$\mathbb{F}(b) \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} I(b_{ij}) + \langle b, \theta\rangle$$

$\rho(i,j) = 0$ Obviously concave upper bound.

$\rho(i,j) = 1$ Bethe free energy.

# Tree-reweighted free energy

$$\mathbb{F}_{\rm TRW}(b) = \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} \rho(ij) I(b_{ij}) + \langle b, \theta \rangle$$

Compare with Bethe free energy

$$\mathbb{F}(b) \sum_{i\in V} H(b_i) - \sum_{(i,j)\in V} I(b_{ij}) + \langle b, \theta \rangle$$

$\rho(i,j) = 0$ Obviously concave upper bound.

$\rho(i,j) = 1$ Bethe free energy.

# Edge weights

$\rho = (\,\rho(e)\;:\; e \in E)$
Intepretation

$$\rho(e) = \mathbb{P}_\rho\{e \in E(T)\}\,, \qquad \mathbb{P}_\rho(T) = \rho_T\,.$$

Spanning-Tree polytope

$$\sum_{(i,j)\in E} \rho(i,j) \;=\; |V| - 1\,,$$
$$\sum_{(i,j)\in E(U)} \rho(i,j) \;\leq\; |U| - 1\,, \qquad \text{for all } U \subseteq V\,.$$

# Edge weights

$\rho = (\, \rho(e) \; : \; e \in E)$
Intepretation

$$\rho(e) = \mathbb{P}_{\rho}\{e \in E(T)\}\,, \qquad \mathbb{P}_{\rho}(T) = \rho_T\,.$$

Spanning-Tree polytope

$$\begin{aligned} \sum_{(i,j)\in E} \rho(i,j) &= |V| - 1\,, \\ \sum_{(i,j)\in E(U)} \rho(i,j) &\leq |U| - 1\,, \qquad \text{for all } U \subseteq V\,. \end{aligned}$$

# Example

$k$-regular graph

$$|V| = n, \qquad |E| = \frac{nk}{2}\,.$$

Take all the weights equal (not necessarily ok, but. . . )

$$\rho(i,j) = \frac{2(n-1)}{nk} \approx \frac{2}{k}$$

For (some) models on locally tree-like graphs, $\rho(i,j) = 1$ is approximately correct $\to \Theta(n)$ error.

# Example

$k$-regular graph

$$|V| = n, \qquad |E| = \frac{nk}{2}\,.$$

Take all the weights equal (not necessarily ok, but. . . )

$$\rho(i,j) = \frac{2(n-1)}{nk} \approx \frac{2}{k}$$

For (some) models on locally tree-like graphs, $\rho(i,j) = 1$ is approximately correct $\rightarrow \Theta(n)$ error.