# Probabilistic Graphical Models

Brown University CSCI 2950-P, Spring 2013

Prof. Erik Sudderth

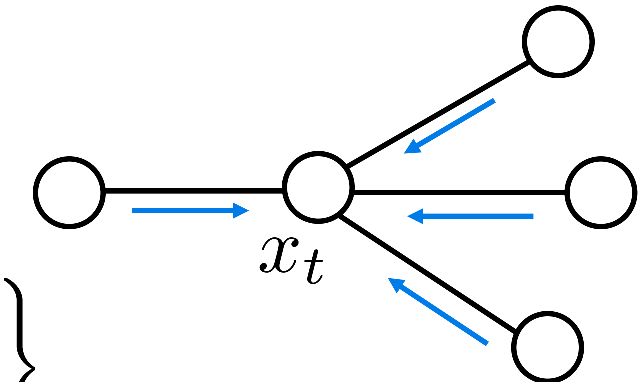
Lecture 20:

Structured Variational Methods, Bethe Approximations and Loopy BP

# Mean Field versus Belief Propagation

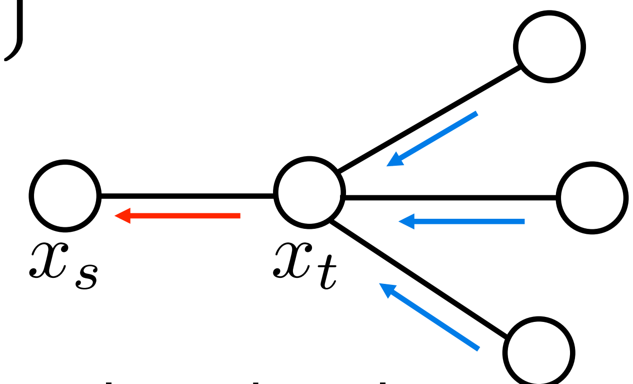
$$p(x) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s) \qquad \phi_{st}(x_s, x_t) = -\psi_{st}(x_s, x_t)$$

$$q_t(x_t) \propto \psi_t(x_t) \prod_{u\in\Gamma(t)} m_{ut}(x_t)$$




**MF:** $m_{ts}(x_s) \propto \exp\left\{ -\sum_{x_t} \phi_{st}(x_s, x_t) q_t(x_t) \right\}$

**BP:** $m_{ts}(x_s) \propto \sum_{x_t} \psi_{st}(x_s, x_t) \frac{q_t(x_t)}{m_{st}(x_t)}$




**Big implications from small changes:**

- Mean Field: Guaranteed to converge for general graphs, always lower-bounds partition function, but approximate even on trees
- Belief Propagation: Produces exact marginals for any tree, but for general graphs no guarantees of convergence or accuracy
- Goal: Can we justify and generalize loopy BP?

# Mean Field Free Energy

$$p(x) = \frac{1}{Z} \exp \left\{ - \sum_{(s,t) \in \mathcal{E}} \phi_{st}(x_s, x_t) - \sum_{s \in \mathcal{V}} \phi_s(x_s) \right\}$$

$$q(x) = \prod_{s \in \mathcal{V}} q_s(x_s)$$

$$\phi_{st}(x_s, x_t) = -\log \psi_{st}(x_s, x_t)$$
$$\phi_s(x_s) = -\log \psi_s(x_s)$$

$$D(q \,||\, p) = -H(q) + \sum_x q(x) E(x) + \log Z$$

Mean Field Entropy:

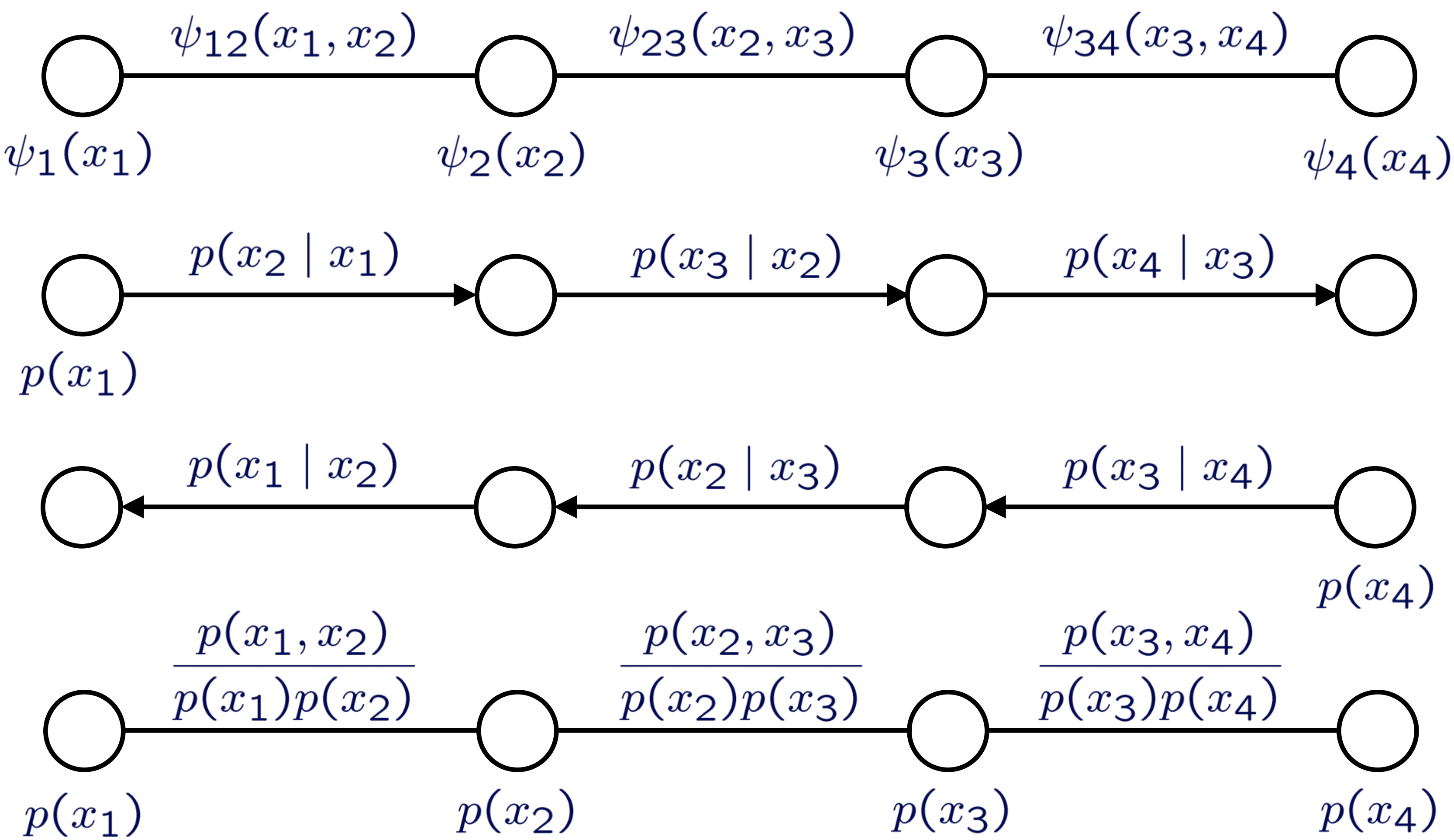
$$H(q) = \sum_{s \in \mathcal{V}} H_s(q_s) = -\sum_{s \in \mathcal{V}} \sum_{x_s} q_s(x_s) \log q_s(x_s)$$

Mean Field Average Energy (expected sufficient statistics):

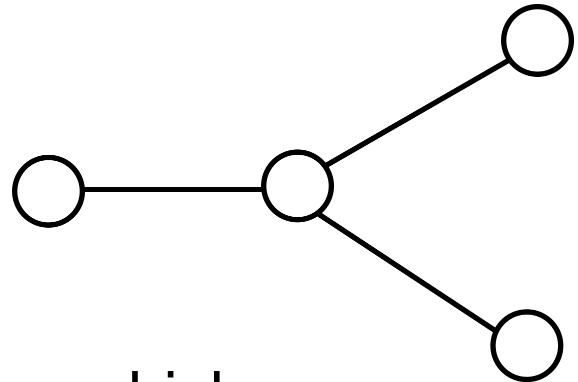
$$\sum_x q(x) E(x) = \sum_{(s,t) \in \mathcal{E}} \sum_{x_s, x_t} q_s(x_s) q_t(x_t) \phi_{st}(x_s, x_t) + \sum_{s \in \mathcal{V}} \sum_{x_s} q_s(x_s) \phi_s(x_s)$$

# Markov Chain Factorizations

$$p(x) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s)$$



# Tree Structured Variational Methods



- Trees exactly factorize as

$$q(x) = \prod_{(s,t)\in\mathcal{E}} \frac{q_{st}(x_s, x_t)}{q_s(x_s), q_t(x_t)} \prod_{s\in\mathcal{V}} q_s(x_s)$$

- We may then optimize over all distributions which are Markov with respect to a tree-structured graph:

$$D(q \,||\, p) = -H(q) + \sum_x q(x)E(x) + \log Z$$

$$\sum_x q(x)E(x) = \sum_{(s,t)\in\mathcal{E}} \sum_{x_s,x_t} q_{st}(x_s, x_t)\phi_{st}(x_s, x_t) + \sum_{s\in\mathcal{V}} \sum_{x_s} q_s(x_s)\phi_s(x_s)$$

$$H(q) = \sum_{s\in\mathcal{V}} H_s(q_s) - \sum_{(s,t)\in\mathcal{E}} I_{st}(q_{st})$$

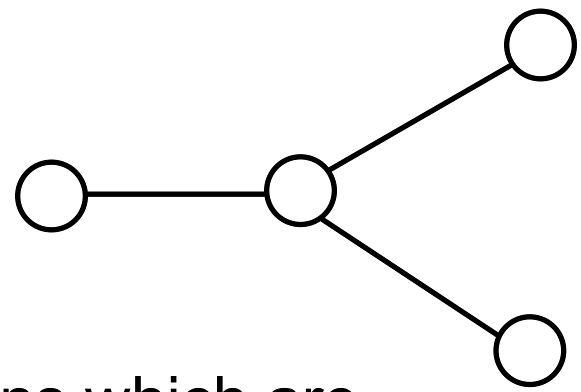
*Marginal Entropies*

*Mutual Information*

# Tree Structured Variational Methods

- Trees exactly factorize as

$$q(x) = \prod_{(s,t)\in\mathcal{E}} \frac{q_{st}(x_s, x_t)}{q_s(x_s), q_t(x_t)} \prod_{s\in\mathcal{V}} q_s(x_s)$$



- We may then optimize over all distributions which are Markov with respect to a tree-structured graph:

$$D(q \,||\, p) = -H(q) + \sum_x q(x)E(x) + \log Z$$

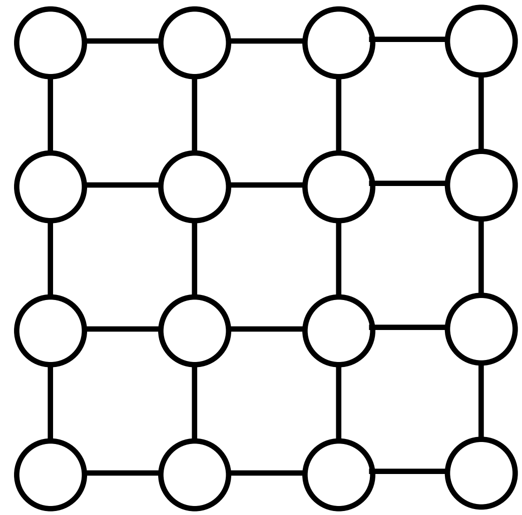
$$\sum_x q(x)E(x) = \sum_{(s,t)\in\mathcal{E}} \sum_{x_s,x_t} q_{st}(x_s, x_t)\phi_{st}(x_s, x_t) + \sum_{s\in\mathcal{V}} \sum_{x_s} q_s(x_s)\phi_s(x_s)$$

$$H(q) = \sum_{s\in\mathcal{V}} H_s(q_s) - \sum_{(s,t)\in\mathcal{E}} I_{st}(q_{st})$$

$$H_s(q_s) = -\sum_{x_s} q_s(x_s)\log q_s(x_s) \qquad I_{st}(q_{st}) = \sum_{x_s,x_t} q_{st}(x_s, x_t)\log \frac{q_{st}(x_s, x_t)}{q_s(x_s)q_t(x_t)}$$

# Bethe Variational Approximations

*Bethe approximation uses the tree-structured free energy form even though the graph has cycles*



$$D(q \,||\, p) = -H(q) + \sum_x q(x)E(x) + \log Z$$

**Average Energy**

$$\sum_x q(x)E(x) = \sum_{(s,t)\in\mathcal{E}} \sum_{x_s,x_t} q_{st}(x_s,x_t)\phi_{st}(x_s,x_t) + \sum_{s\in\mathcal{V}} \sum_{x_s} q_s(x_s)\phi_s(x_s)$$

**Approximate Entropy**

$$H(q) \approx \sum_{s\in\mathcal{V}} H_s(q_s) - \sum_{(s,t)\in\mathcal{E}} I_{st}(q_{st})$$

$$\tilde{q}_s(x_s) = \sum_{x_t} q_{st}(x_s,x_t)$$

$$H_s(q_s) = -\sum_{x_s} q_s(x_s)\log q_s(x_s)$$

$$I_{st}(q_{st}) = \sum_{x_s,x_t} q_{st}(x_s,x_t)\log \frac{q_{st}(x_s,x_t)}{\tilde{q}_s(x_s)\tilde{q}_t(x_t)}$$

*Optimization must enforce marginalization constraints*

# Bethe Variational Lagrangian

$$\mathcal{L}(q,\lambda) =$$

$$+\sum_{s\in\mathcal{V}}\sum_{x_s} q_s(x_s)(\phi_s(x_s) + \log q_s(x_s))$$

$$\tilde{q}_s(x_s) = \sum_{x_t} q_{st}(x_s,x_t)$$

$$+\sum_{(s,t)\in\mathcal{E}}\sum_{x_s,x_t} q_{st}(x_s,x_t)\left(\phi_{st}(x_s,x_t) + \log\frac{q_{st}(x_s,x_t)}{\tilde{q}_s(x_s)\tilde{q}_t(x_t)}\right)$$

$$+\sum_{s\in\mathcal{V}}\lambda_{ss}\left(1-\sum_{x_s} q_s(x_s)\right)$$

$$+\sum_{(s,t)\in\mathcal{E}}\left[\sum_{x_s}\lambda_{ts}(x_s)\left(q_s(x_s)-\sum_{x_t} q_{st}(x_s,x_t)\right) + \sum_{x_t}\lambda_{st}(x_t)\left(q_t(x_t)-\sum_{x_s} q_{st}(x_s,x_t)\right)\right]$$

*Constraints not explicitly enforced:*

$$1-\sum_{x_s,x_t} q_{st}(x_s,x_t) = 0$$

*Implied by other equality constraints*

$$q_s(x_s)\geq 0,\, q_{st}(x_s,x_t)\geq 0$$

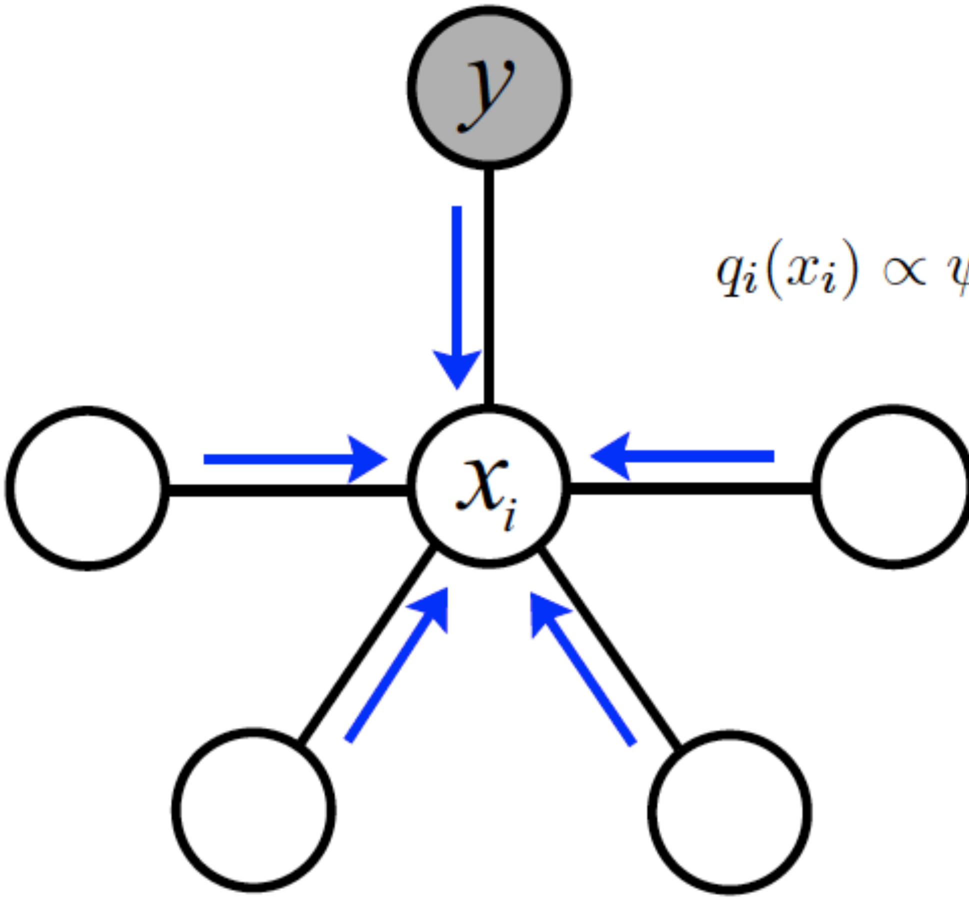
*Inactive, will be automatically satisfied by solution we derive*

# Derivation: Bethe to Loopy BP

Derivation on whiteboard. For details, see:

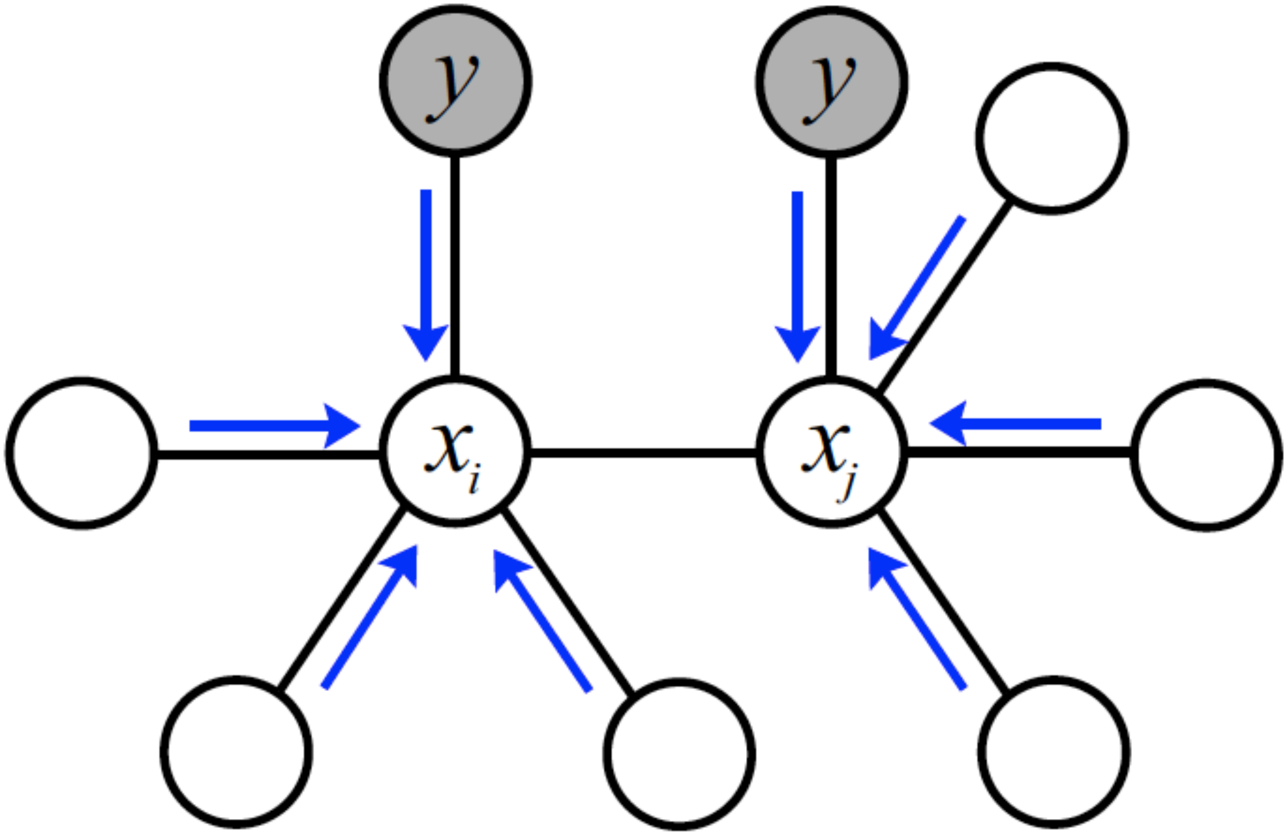
- Wainwright & Jordan, *Graphical Models, Exponential Families, & Variational Inference*.
  Foundations and Trends in Machine Learning, 2008, Sec. 4.1.
- Yedidia, Freeman, & Weiss, *Understanding Belief Propagation and its Generalizations*.
  Exploring Artificial Intelligence in the New Millennium, 2002.

# BP Algorithm



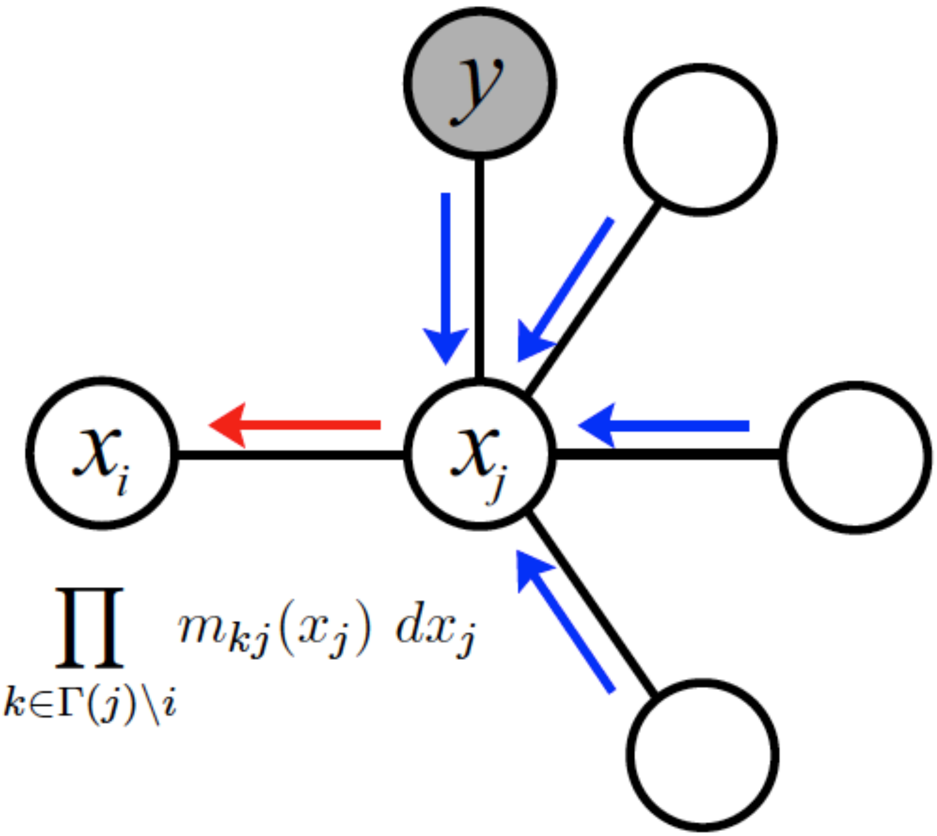

$$q_i(x_i) \propto \psi_i(x_i, y) \prod_{j \in \Gamma(i)} m_{ji}(x_i)$$

# BP Algorithm




$$q_{ij}(x_i, x_j) \propto \psi_{ij}(x_i, x_j)\, \psi_i(x_i, y)\, \psi_j(x_j, y) \prod_{\ell \in \Gamma(i) \backslash j} m_{\ell i}(x_i) \prod_{k \in \Gamma(j) \backslash i} m_{kj}(x_j)$$

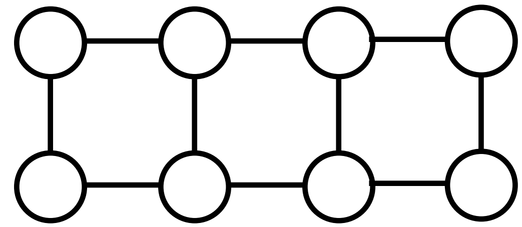
# BP Algorithm




$$m_{ji}(x_i) \propto \int_{\mathcal{X}_j} \psi_{ij}(x_i, x_j)\, \psi_j(x_j, y) \prod_{k \in \Gamma(j) \setminus i} m_{kj}(x_j)\; dx_j$$

# Bethe Approximations and Loopy BP

$$D(q \,||\, p) = -H(q) + \sum_x q(x)E(x) + \log Z$$



$$\sum_x q(x)E(x) = \sum_{(s,t)\in\mathcal{E}} \sum_{x_s,x_t} q_{st}(x_s,x_t)\phi_{st}(x_s,x_t) + \sum_{s\in\mathcal{V}} \sum_{x_s} q_s(x_s)\phi_s(x_s)$$

$$H(q) \approx \sum_{s\in\mathcal{V}} H_s(q_s) - \sum_{(s,t)\in\mathcal{E}} I_{st}(q_{st})$$

- For a tree-structure graphical model, entropy approximation becomes exact, and unique solution gives true marginals
- For general graphs, there is a correspondence between *fixed points* of the loopy belief propagation algorithm and *stationary points* of the Bethe variational objective
- Biggest practical applications:
  - Alternative, stable algorithms for Bethe objective
  - Message passing algorithms from fancier objectives

# Implications for Loopy BP

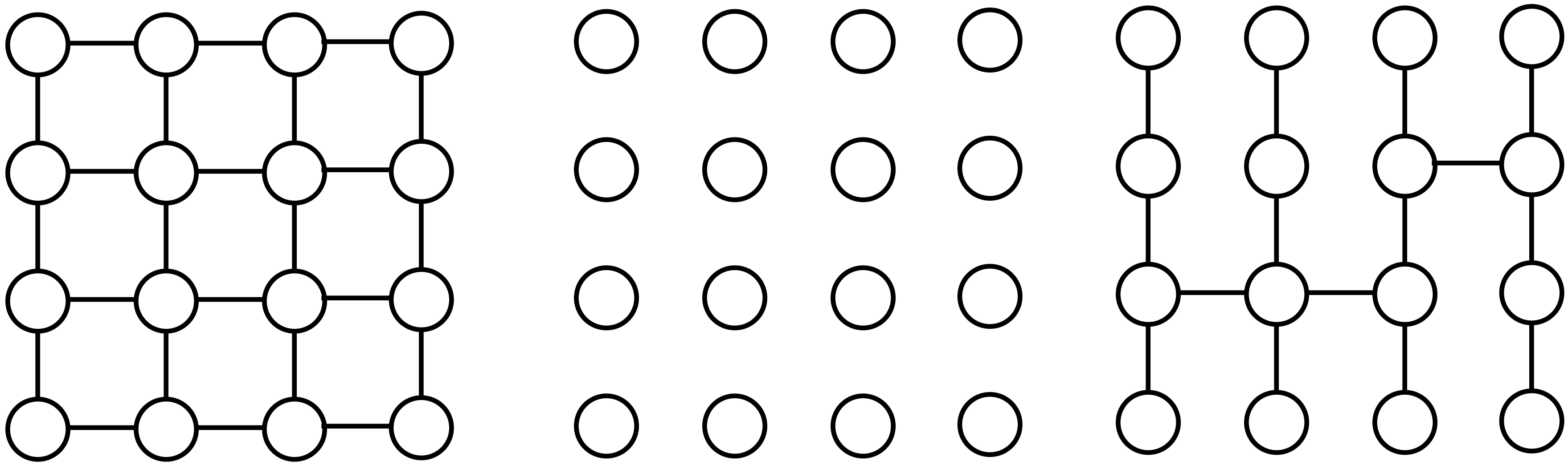
### Bethe Free Energy is an Approximation

- BP may have multiple fixed points (non-convex)
- BP is not guaranteed to converge
- Few general guarantees on BP's accuracy

### Characterizations of BP Fixed Points

- All graphical models have at least one BP fixed point
- Stable fixed points are local minima of Bethe
- For graphs with cycles, BP is almost never exact
- As cycles grow long, BP becomes exact (coding)

# Structured Mean Field



**Original Graph** **Naïve Mean Field** **Structured Mean Field**

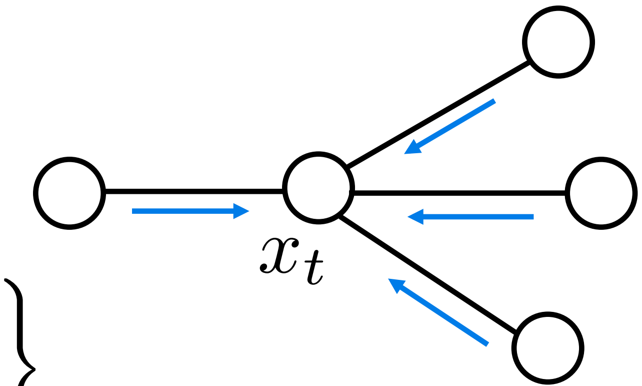
- Any subgraph for which inference is tractable leads to a mean field style approximation for which the update equations are tractable

# Structured Mean Field

$$p(x) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s)$$

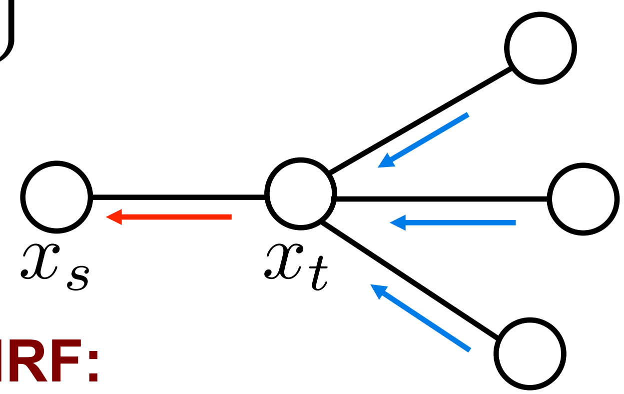
$$\phi_{st}(x_s, x_t) = -\psi_{st}(x_s, x_t)$$

$$q_t(x_t) \propto \psi_t(x_t) \prod_{u\in\Gamma(t)} m_{ut}(x_t)$$




**MF:** $$m_{ts}(x_s) \propto \exp\left\{ -\sum_{x_t} \phi_{st}(x_s, x_t) q_t(x_t) \right\}$$

**BP:** $$m_{ts}(x_s) \propto \sum_{x_t} \psi_{st}(x_s, x_t) \frac{q_t(x_t)}{m_{st}(x_t)}$$




**For the special case of a discrete pairwise MRF:**

- Choose a subset of *core* edges which form no cycles
- On *core* edges, apply BP message updates
- On *other* edges, apply MF message updates
- Guaranteed convergent, optimizes lower bound on Z