CS 3710 Advanced Topics in AI
Lecture 15

Constructing Free Energy Approximations and GBP Algorithms

Branislav Kveton

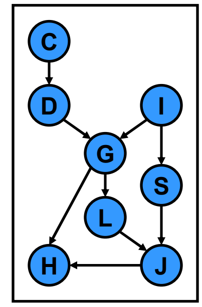
bkveton@cs.pitt.edu

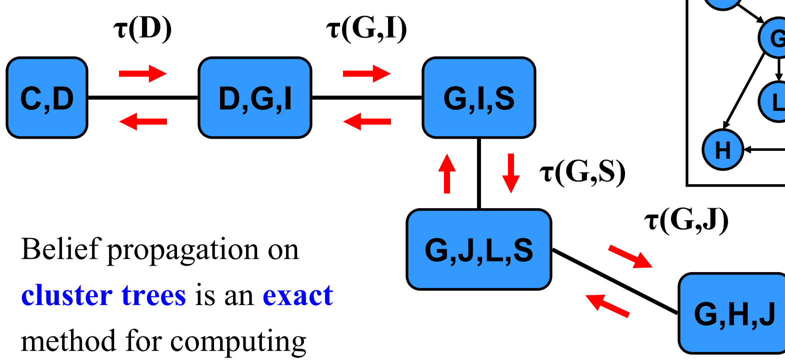
5802 Sennott Square

Content

- Why?
- Belief propagation (BP)
- Factor graphs
- Region-based free energy approximations
- Bethe method
- Bethe method and BP
- Region graphs
- Generalized belief propagation (GBP)

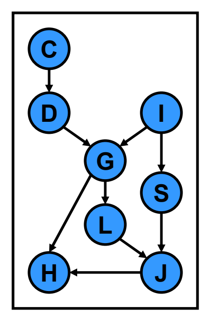
BP on Cluster Trees

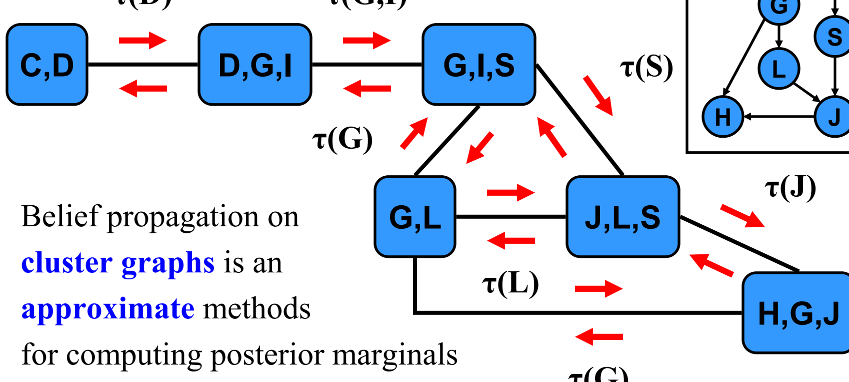




- Belief propagation on **cluster trees** is an **exact** method for computing posterior marginals
- Space complexity is exponential in the **treewidth** of the graph

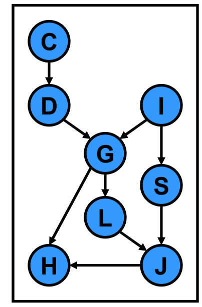
BP on Cluster Graphs

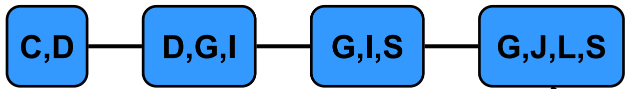


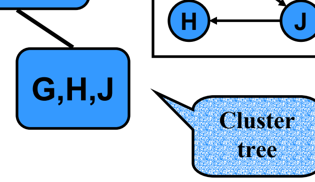


- Belief propagation on **cluster graphs** is an **approximate** methods for computing posterior marginals
- Convergence is **not guaranteed**!
- Space complexity is **linear** in the size of the **largest cluster**

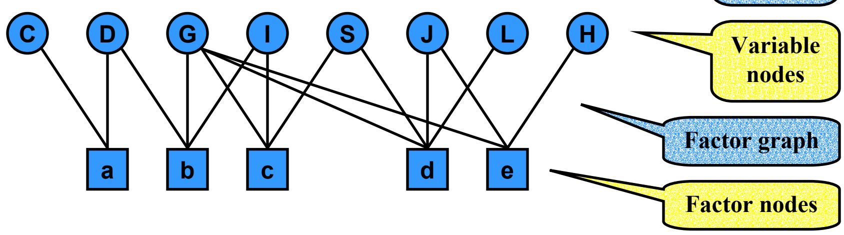
Factor Graphs



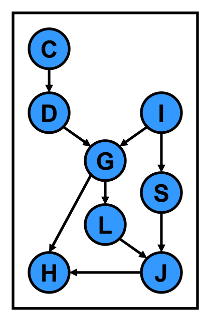


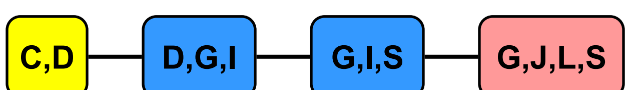


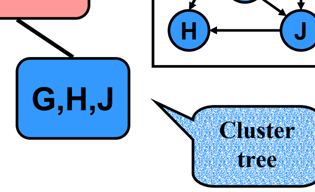
- A **factor graph** is a bipartite graph that represents factored structure



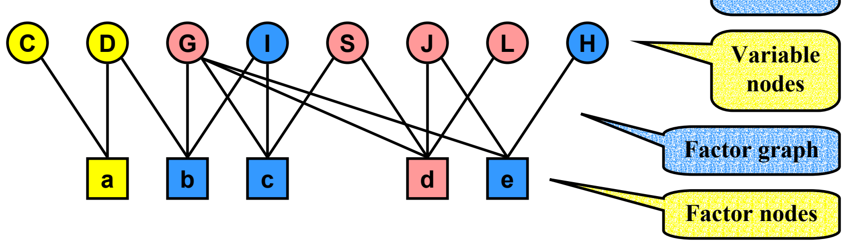
Factor Graphs

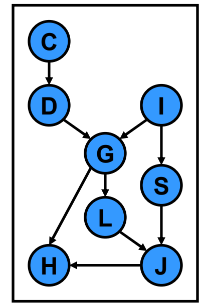






- A **factor graph** is a bipartite graph that represents factored structure





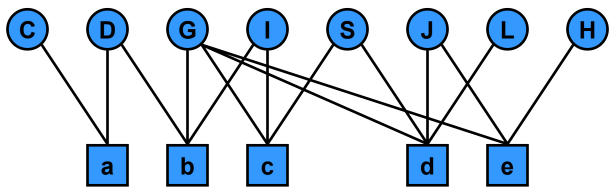
BP on Factor Graphs

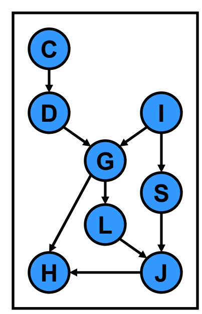
- **Messages**:

$$n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$$

$$m_{a\to i}(x_i) = \sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$$

- **Marginals**: $b(x_i) \propto \prod_{a\in N(i)} m_{a\to i}(x_i)$





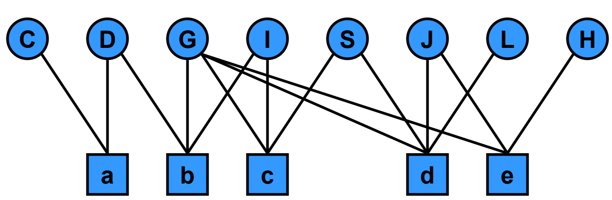
BP on Factor Graphs

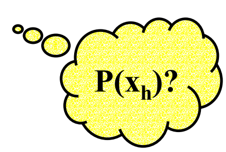
- **Messages**:

$$n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$$

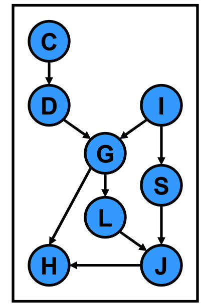
$$m_{a\to i}(x_i) = \sum_{\mathbf{x}_a\setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$$

- **Marginals**: $b(x_i) \propto \prod_{a\in N(i)} m_{a\to i}(x_i)$





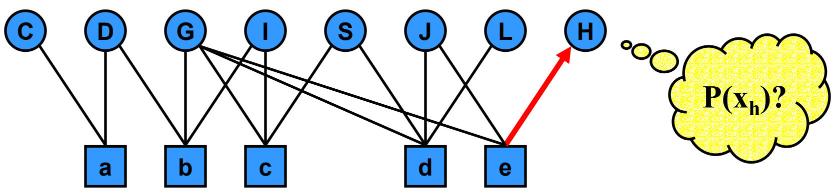
BP on Factor Graphs



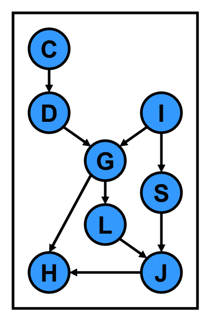
- **Messages**: $n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$

  $m_{a\to i}(x_i) = \sum_{\mathbf{x}_a \setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$

- **Marginals**: $b(x_i) \propto \prod_{a\in N(i)} m_{a\to i}(x_i)$



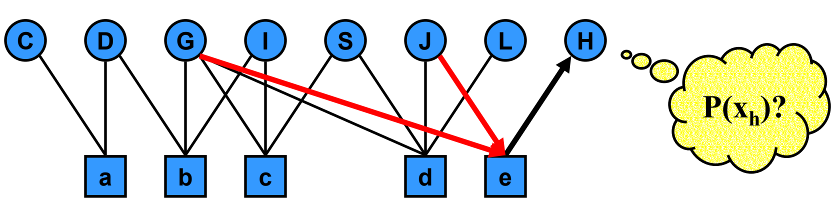
BP on Factor Graphs



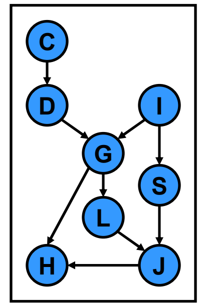
- **Messages**: $n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$

  $m_{a\to i}(x_i) = \sum_{\mathbf{x}_a \setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$

- **Marginals**: $b(x_i) \propto \prod_{a\in N(i)} m_{a\to i}(x_i)$



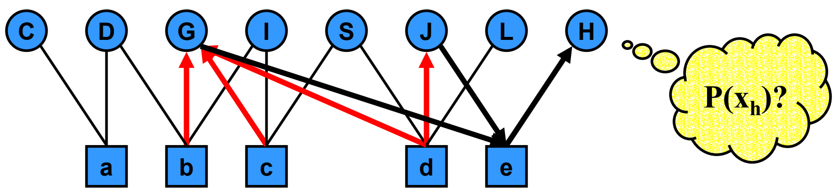
BP on Factor Graphs



- **Messages**: $n_{i\to a}(x_i) = \prod_{c\in N(i)\backslash a} m_{c\to i}(x_i)$

  $m_{a\to i}(x_i) = \sum_{\mathbf{x}_a \backslash x_i} f_a(\mathbf{x}_a) \prod_{j\in N(a)\backslash i} n_{j\to a}(x_j)$

- **Marginals**: $b(x_i) \propto \prod_{a\in N(i)} m_{a\to i}(x_i)$



Free Energy

- **Factored joint**:

$$p(\mathbf{x}) = \exp[-E(\mathbf{x})]/Z \qquad E(\mathbf{x}) = -\sum_{a=1}^{M} \ln f_a(\mathbf{x}_a)$$

$$Z = \sum_{\mathbf{x}\in S} \exp[-E(\mathbf{x})]$$

- **Energy equations**:

$$F(b) = F_H + D(b\|p) \qquad F_H = -\ln Z$$

$$F(b) = U(b) - H(b) \qquad U(b) = \sum_{\mathbf{x}\in S} b(\mathbf{x}) E(\mathbf{x})$$

$$H(b) = -\sum_{\mathbf{x}\in S} b(\mathbf{x}) \ln b(\mathbf{x})$$

Helmholtz free energy

Variational average energy

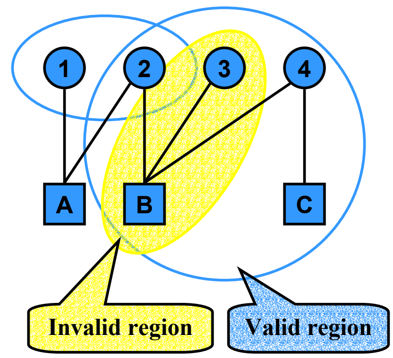
Variational entropy

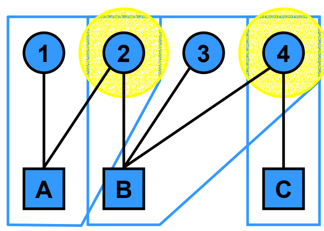
Gibbs free energy

$F(b) \geq F_H$

Region-Based Free Energy

- A **region R** of a factor graph is given by a variable set V_R and a factor set A_R such that if a factor node belongs to A_R, all its variable nodes are in V_R
- The formalism can express both **Kikuchi** and **Bethe** approximations





Region-Based Free Energy

- **Region average energy and entropy**:

$$F_R(b_R) = U_R(b_R) - H_R(b_R) \qquad U_R(b_R) = \sum_{\mathbf{x}_R} b_R(\mathbf{x}_R) E_R(\mathbf{x}_R)$$

$$E_R(\mathbf{x}_R) = -\sum_{a \in A_R} \ln f_a(\mathbf{x}_a) \qquad H_R(b_R) = -\sum_{\mathbf{x}_R} b_R(\mathbf{x}_R) \ln b_R(\mathbf{x}_R)$$

- **Region-based average energy and approximate entropy**:

$$F_{\Re}(\{b_R\}) = U_{\Re}(\{b_R\}) - H_{\Re}(\{b_R\}) \qquad U_{\Re}(\{b_R\}) = \sum_{R \in \Re} c_R U_R(b_R)$$

$$H_{\Re}(\{b_R\}) = \sum_{R \in \Re} c_R H_R(b_R)$$

where c_R is a **counting number** of the region R

Region-Based Free Energy

- A region-based approximation is **valid** if for every factor node a and every variable node i:

$$\sum_{R\in\Re} c_R I_{A_R}(a) = \sum_{R\in\Re} c_R I_{V_R}(i) = 1$$

Neither factors nor variables are double counted

- If $b_R(\mathbf{x}_R) = p_R(\mathbf{x}_R)$, then the average energy $U_\Re$ of a valid region-based approximation is **exact**
- If $p(\mathbf{x})$ is **uniform** and $b_R(\mathbf{x}_R) = p_R(\mathbf{x}_R)$, the entropy $H_\Re$ of a valid region-based approximation is **exact**

Minimization of $F_\Re$ can be achieved by minimizing $D(b||p)$ or maximizing $H_\Re$

Constrained Region-Based Free Energy

- A region-based approximation is **constrained** if:
  - Every $b_R(\mathbf{x}_R)$ has the form of a probability function
  - Marginals are consistent across regions
- A constrained region-based approximation is **maxent-normal** if it is valid and the entropy $H_\Re$ achieves its maximum when all $b_R(\mathbf{x}_R)$ are **uniform**
- Despite these restrictions we may get **strange looking** results!

Minimization of $F_\Re$ can be achieved by maximizing $H_\Re$

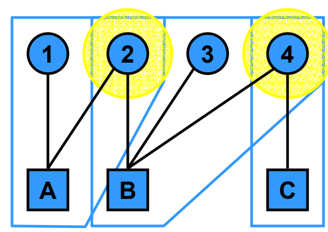
Bethe Method

- **Bethe approximation** is a special case of region-based free energy approximation
- **Bethe free energy** equals to $F_{\text{Bethe}} = U_{\text{Bethe}} - H_{\text{Bethe}}$:

$$U_{\text{Bethe}}(b) = -\sum_{a=1}^{M}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln f_a(\mathbf{x}_a)$$

$$H_{\text{Bethe}}(b) = -\sum_{a=1}^{M}\sum_{\mathbf{x}_a} b_a(\mathbf{x}_a)\ln b_a(\mathbf{x}_a) + \sum_{i=1}^{N}(d_i - 1)\sum_{x_i} b_i(x_i)\ln b_i(x_i)$$

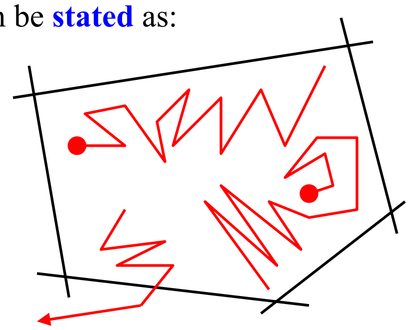
- If the factor graph has **no cycles**, U_{Bethe} and H_{Bethe} are **exact**
- Bethe approximations are **maxent-normal**



Bethe Method and BP

- The problem of finding F_{Bethe} can be **stated** as:

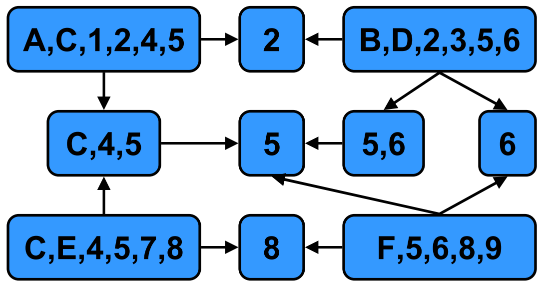
$$\begin{aligned}
&\text{minimize} && F_{\text{Bethe}} \\
&\text{subject to:} && \sum_{\mathbf{x}_a} b_a(\mathbf{x}_a) = 1 \\
& && \sum_{x_i} b_i(x_i) = 1 \\
& && \sum_{\mathbf{x}_a \setminus x_i} b_a(\mathbf{x}_a) = b_i(x_i)
\end{aligned}$$



- To solve this **constrained** optimization problem, we introduce Lagrangian multipliers, and turn it into **unconstrained** one
- The **first derivatives** with respect to beliefs yield **stationary points**, which are **fixed points** of the BP algorithm

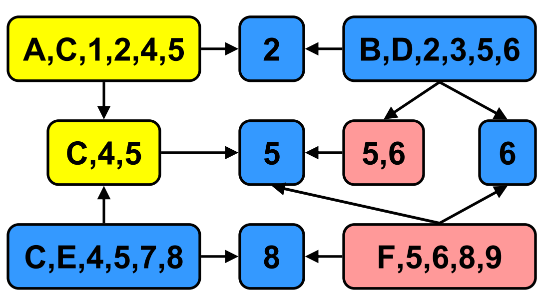
Region Graphs

- **Region graph** is a graphical formalism for generating region-based free energy approximations



Region Graphs

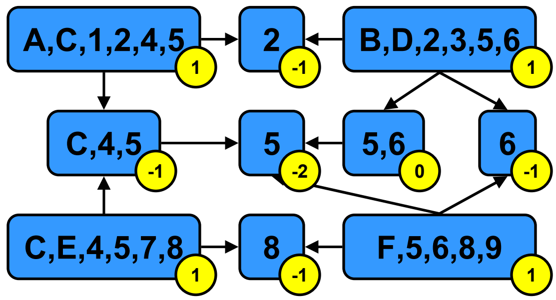
- **Region graph** is a graphical formalism for generating region-based free energy approximations



- Regions R_b and R_e can be **connected** by an edge $R_b \rightarrow R_e$ only if $R_e \subseteq R_b$

Region Graphs

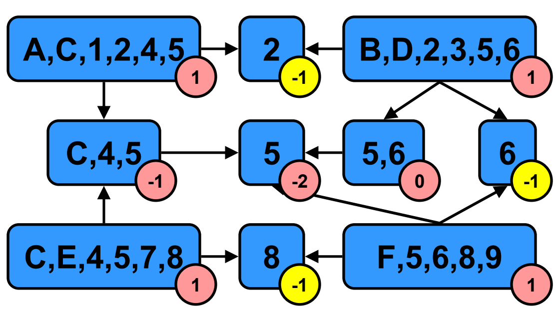
- **Region graph** is a graphical formalism for generating region-based free energy approximations



- The **counting number** equals to one minus the in-degree

Region Graphs

- **Region graph** is a graphical formalism for generating region-based free energy approximations



- The graph has to satisfy the **region graph condition**

Generalized Belief Propagation

- A class of message-passing algorithms
- **Parent-to-child algorithm**
  - Generalizes the BP algorithm on **region graphs**

$$m_{P\to R}(x_R) = \frac{\sum_{x_{P\setminus R}} \prod_{a\in F_{P\setminus R}} f_a(x_a) \prod_{(I,J)\in N(P,R)} m_{I\to J}(x_J)}{\prod_{(I,J)\in D(P,R)} m_{I\to J}(x_J)}$$

$$b_R(\mathbf{x}_R) \propto \prod_{a\in A_R} f_a(\mathbf{x}_a) \left(\prod_{P\in P(R)} m_{P\to R}(\mathbf{x}_R)\right)\left(\prod_{D\in D(R)} \prod_{P'\in P(D)\setminus \varepsilon(R)} m_{P'\to D}(\mathbf{x}_D)\right)$$

  - **Correctness** can be proved similarly to the BP algorithm