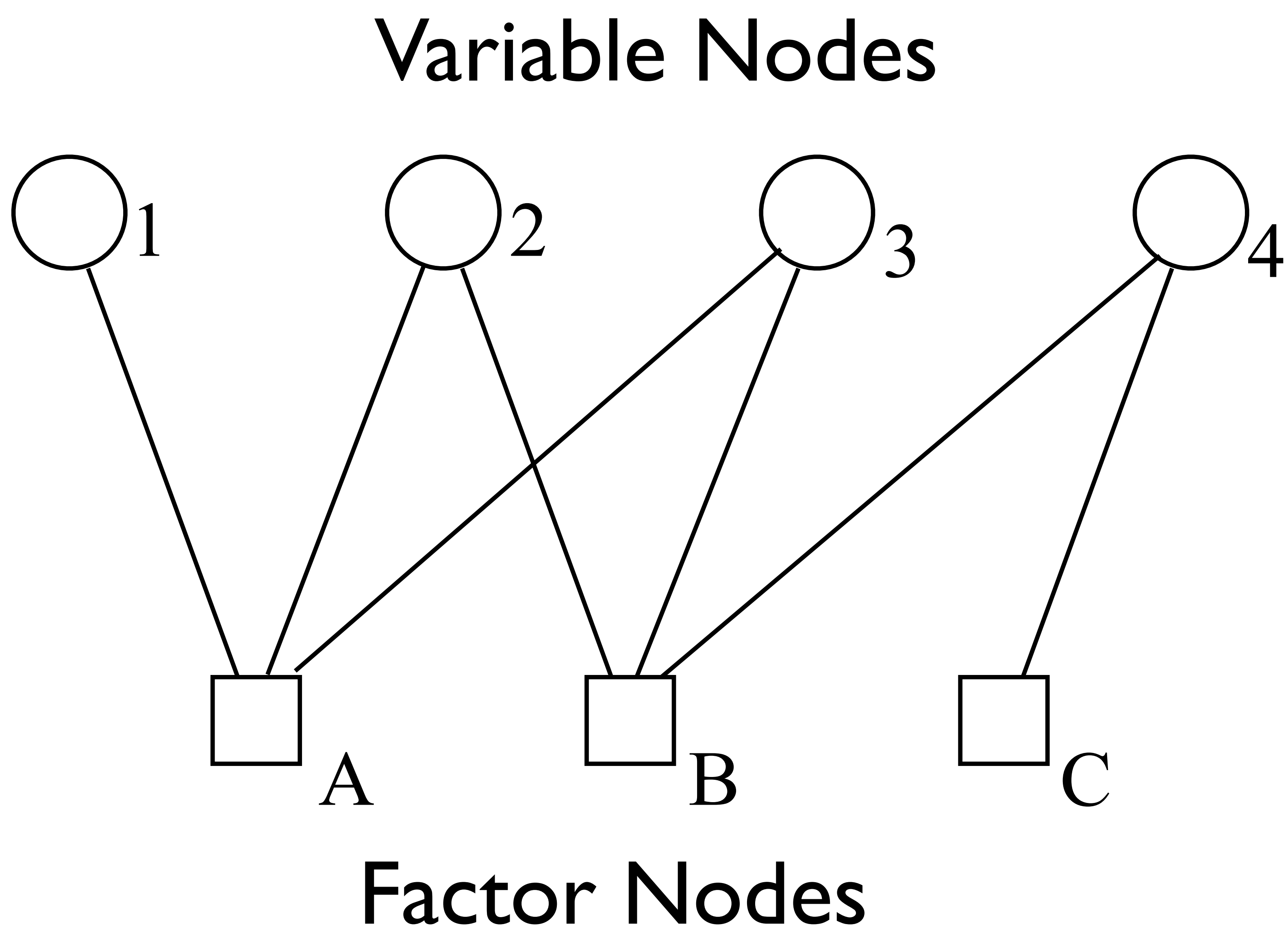
Belief Propagation Tutorial

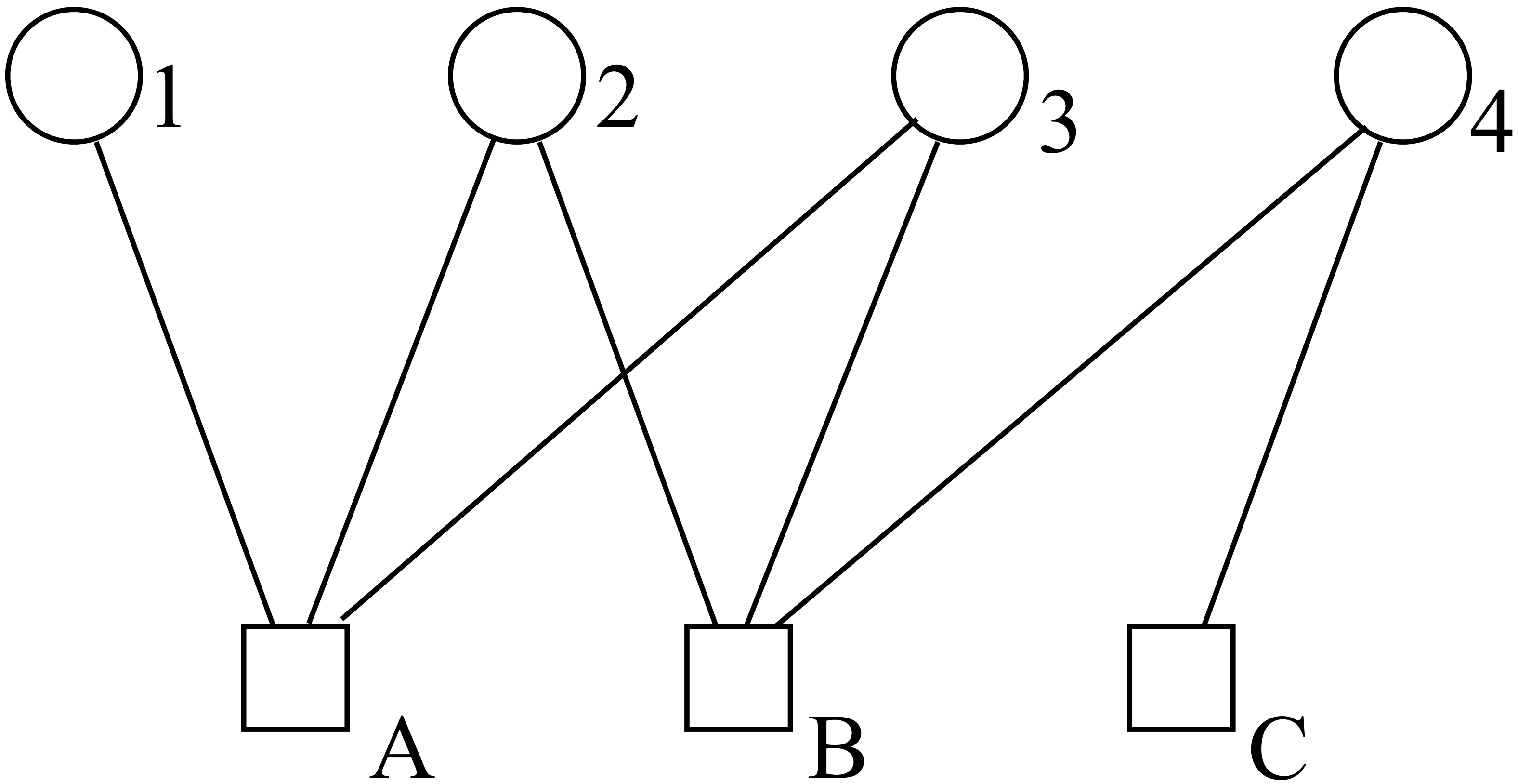
Jonathan Yedidia
Mitsubishi Electric Research Labs

Outline

- Factor Graphs
- Message Passing Algorithms
- Free Energy Approximations

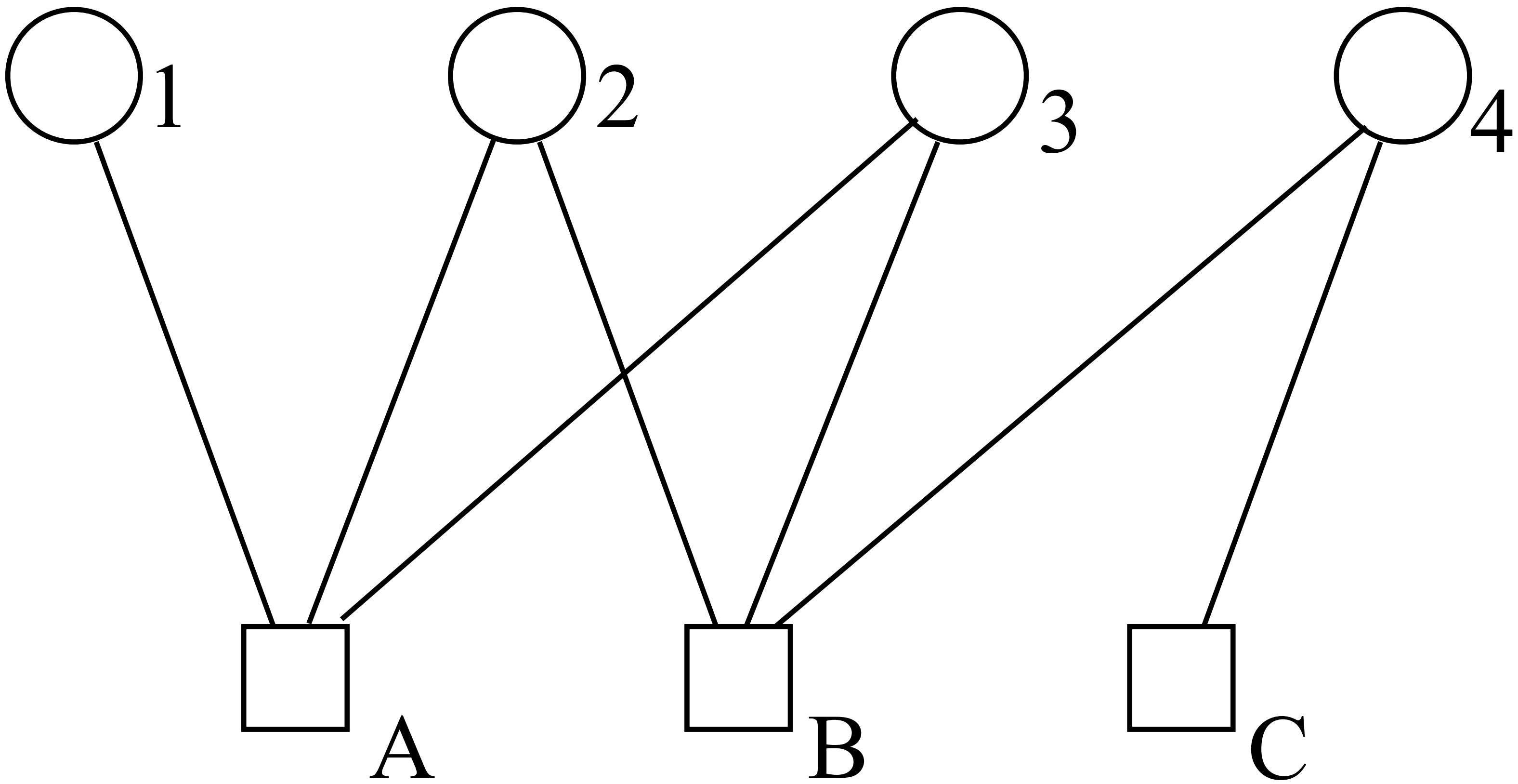

Variable Nodes
1
2
3
4
A
B
C
Factor Nodes

$$p(X) = \frac{1}{Z}\prod_a f_a(X_a)$$



$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} f_A(x_1, x_2, x_3) f_B(x_2, x_3, x_4) f_C(x_4)$$

$$p(X) = \frac{1}{Z} \exp\left(-\sum_a E_a(X_a)\right)$$



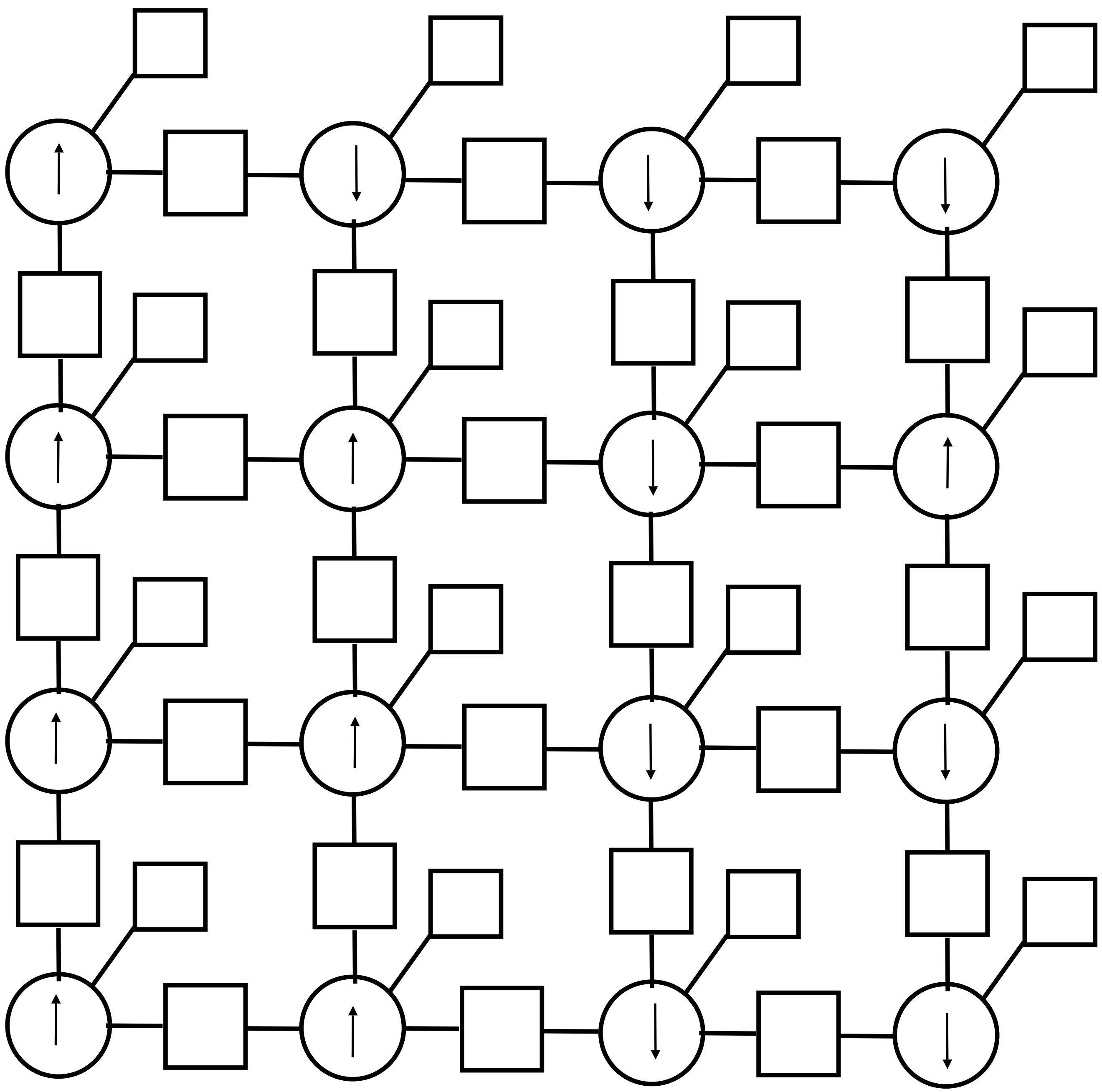
Questions

- What is the most probable configuration?
- What is the marginal probability for a node or group of nodes?

These questions can be attacked by Simulated Annealing or Monte Carlo Simulation, but there are often much more efficient methods.

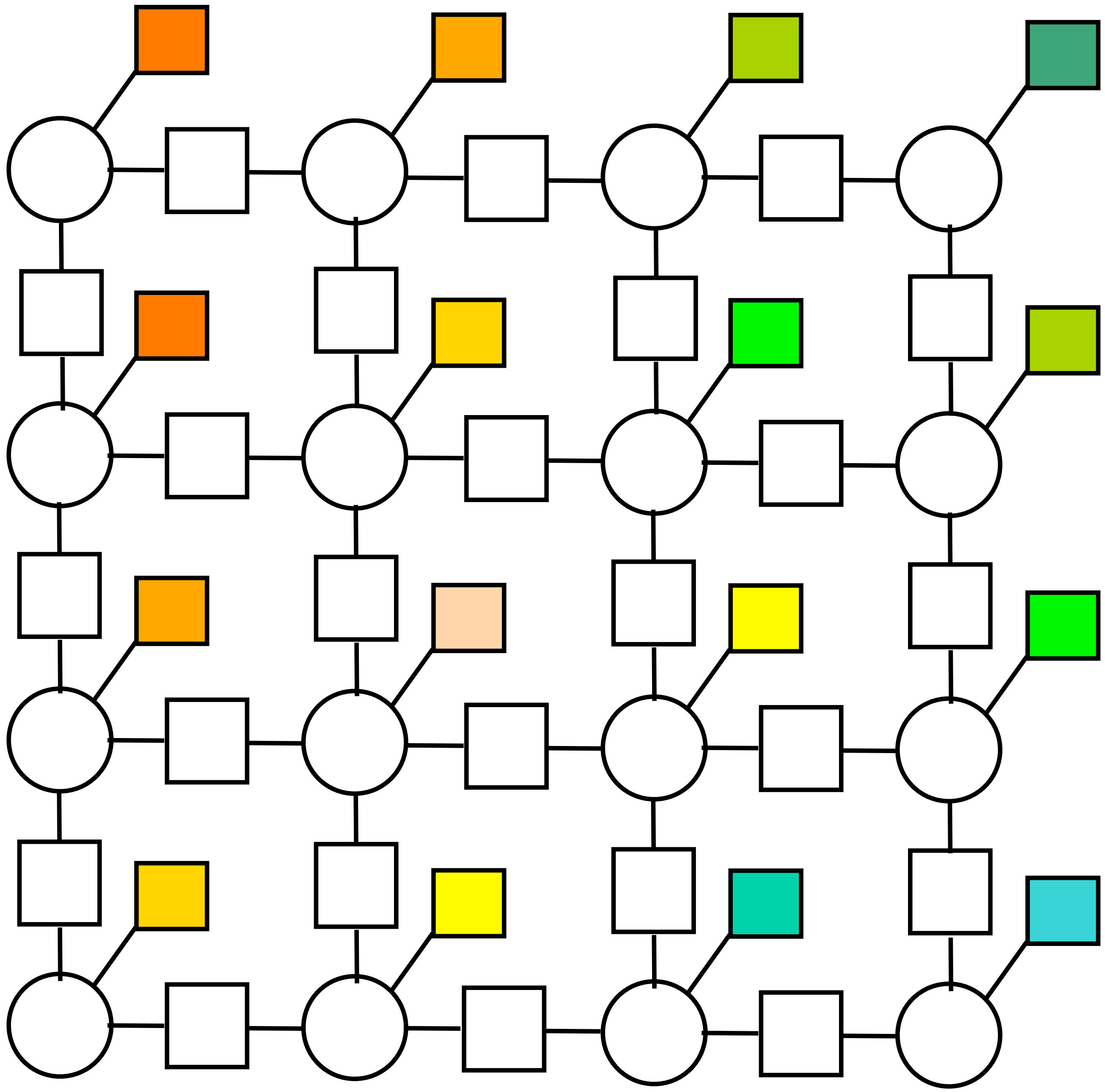
Belief propagation is one method, but one should remember that there are others!

Ising Model



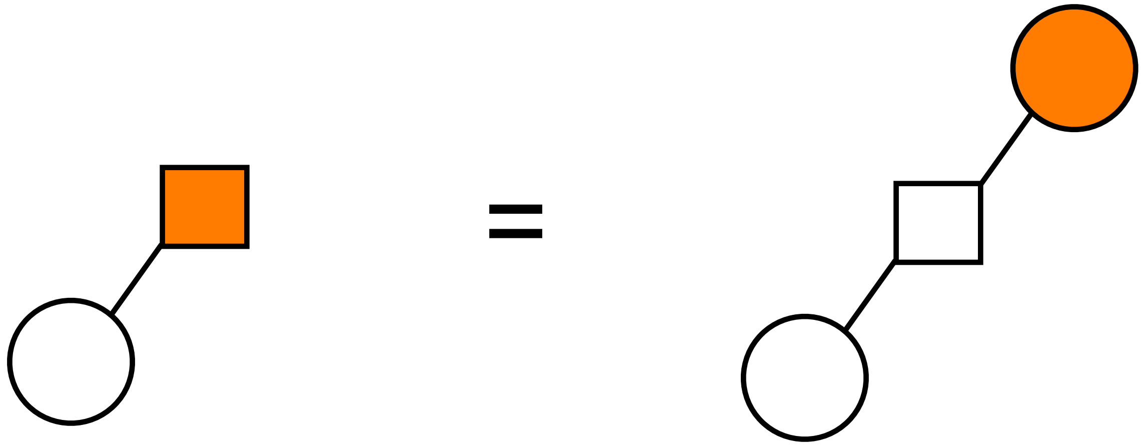
Marginal Probabilities =
local magnetization

Computer Vision and Signal Processing



Marginal Probabilities = "beliefs" about possible underlying scenes

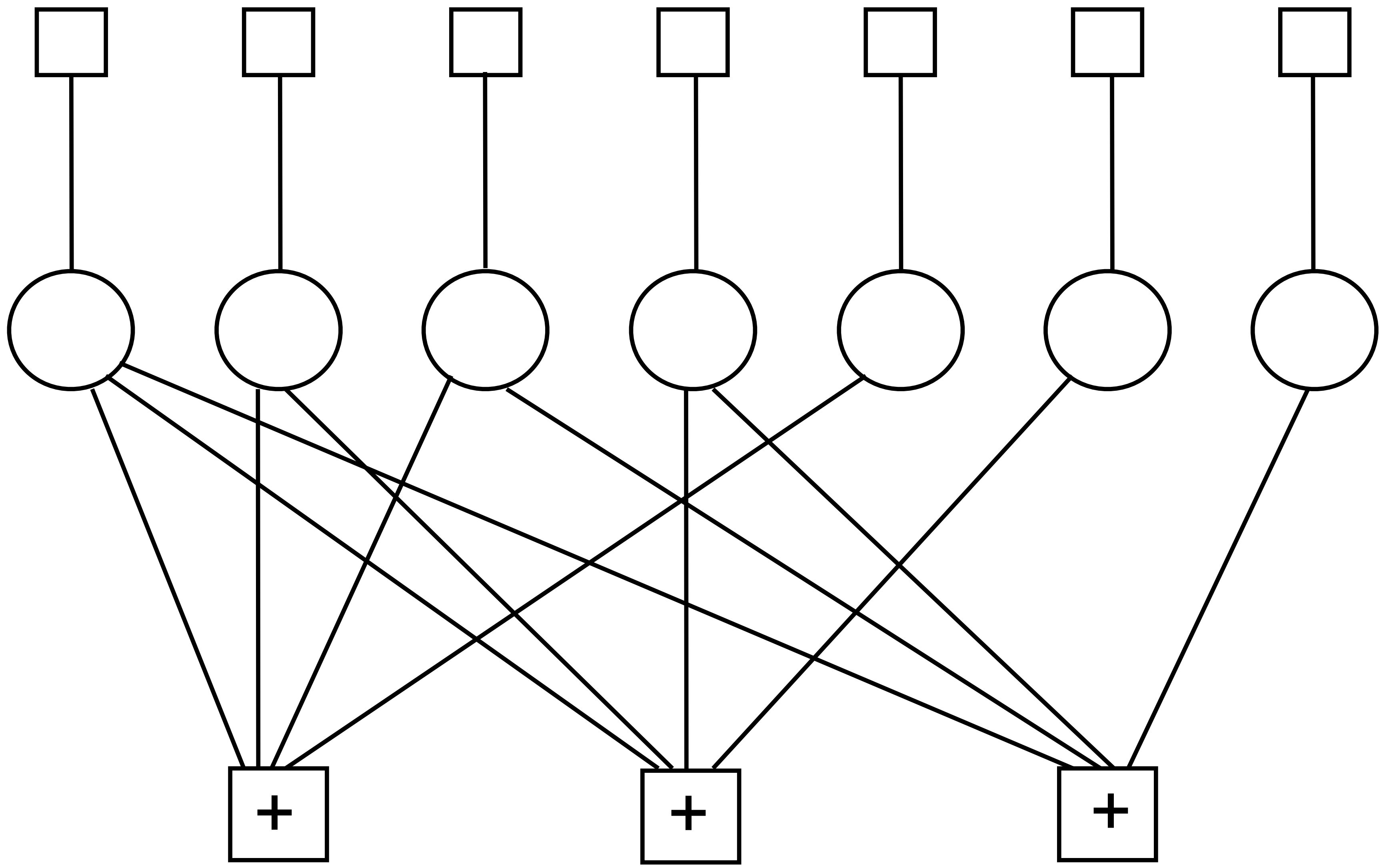
Observed Variables



Observed Variables induce a local "evidence" factor node.

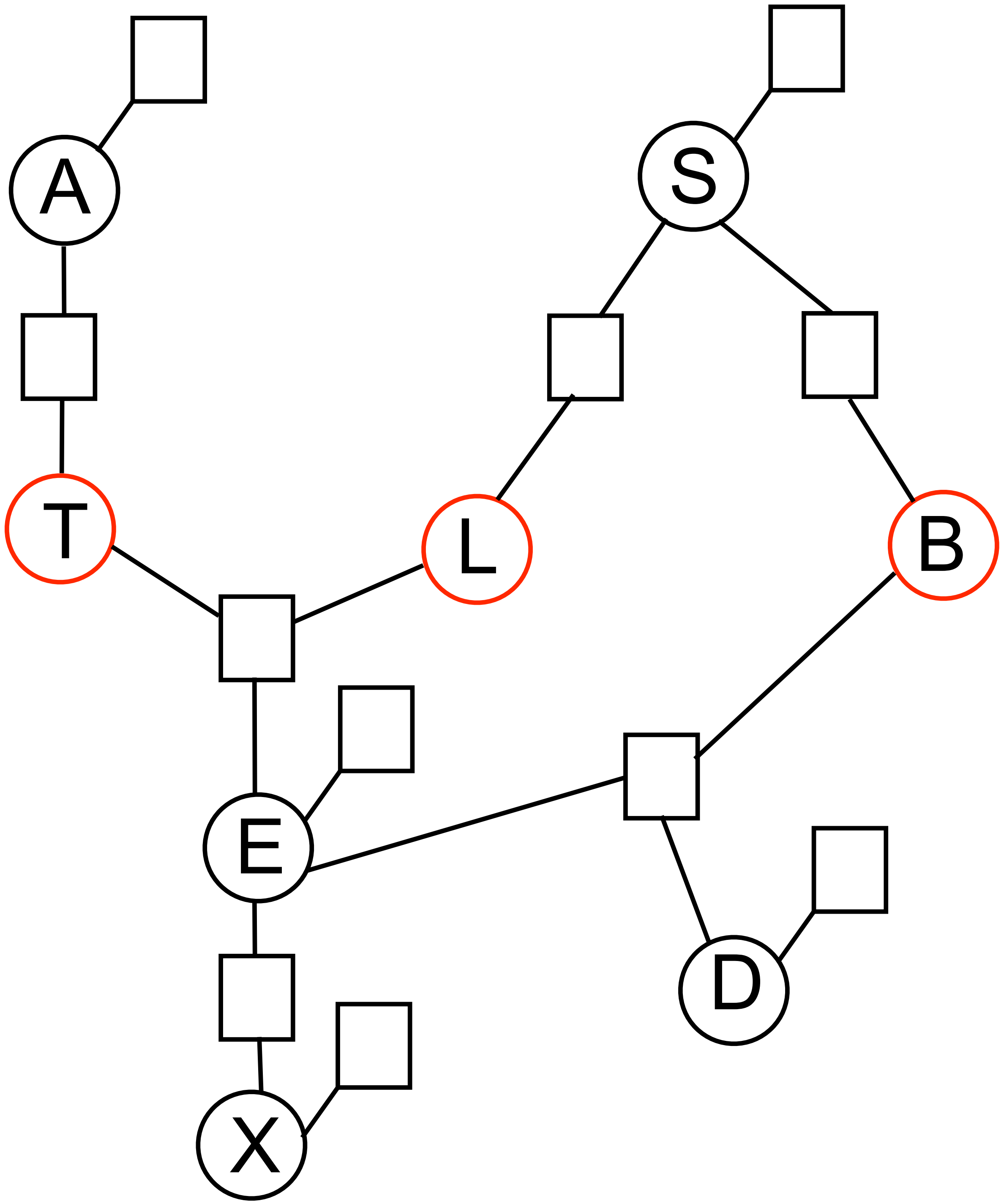
Error-correcting Codes

(Tanner, 1981
Gallager, 1963)



Marginal Probabilities = *A posteriori* bit probabilities

Inference Engines (Pearl, 1988)



Marginal Probabilities = “beliefs” about possible diagnoses

(example adapted from Lauritzen, 1992)

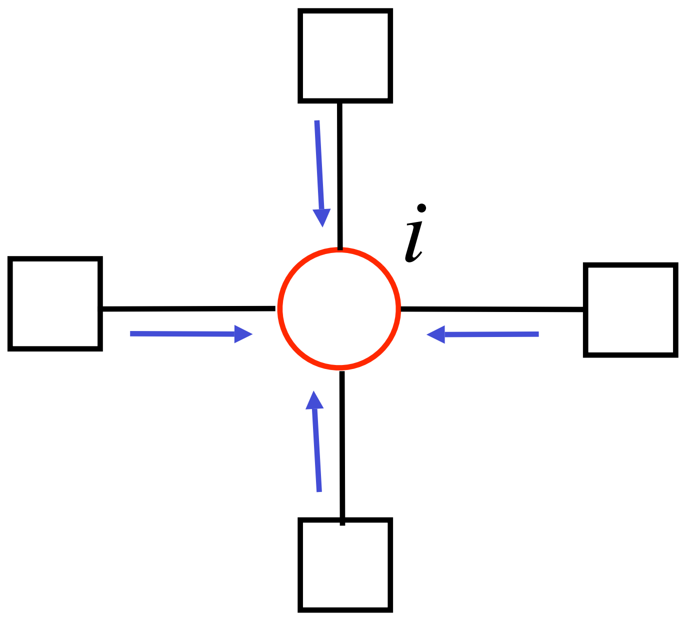
Equivalent Graphical Models

- Markov Random Fields: Use only variable nodes. Factor nodes are defined implicitly when all nodes in a "clique" are connected to each other.
- Forney Factor Graphs: Use only factor nodes. Variables live on edges. Use equality factor nodes to convert ordinary factor graphs into Forney factor graphs
- Bayesian Networks: Use directed graphs which imply conditional probabilities.

Outline

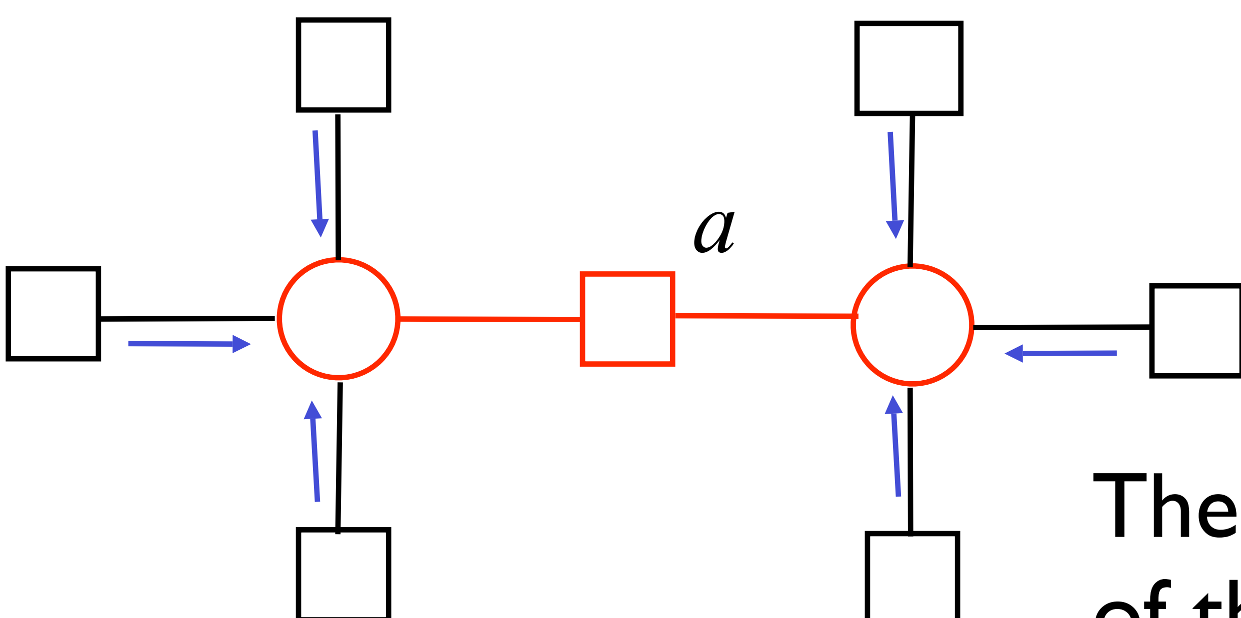
- Factor Graphs
- Message Passing Algorithms
- Free Energy Approximations

Belief Propagation



$$b_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$$

"beliefs" "messages"



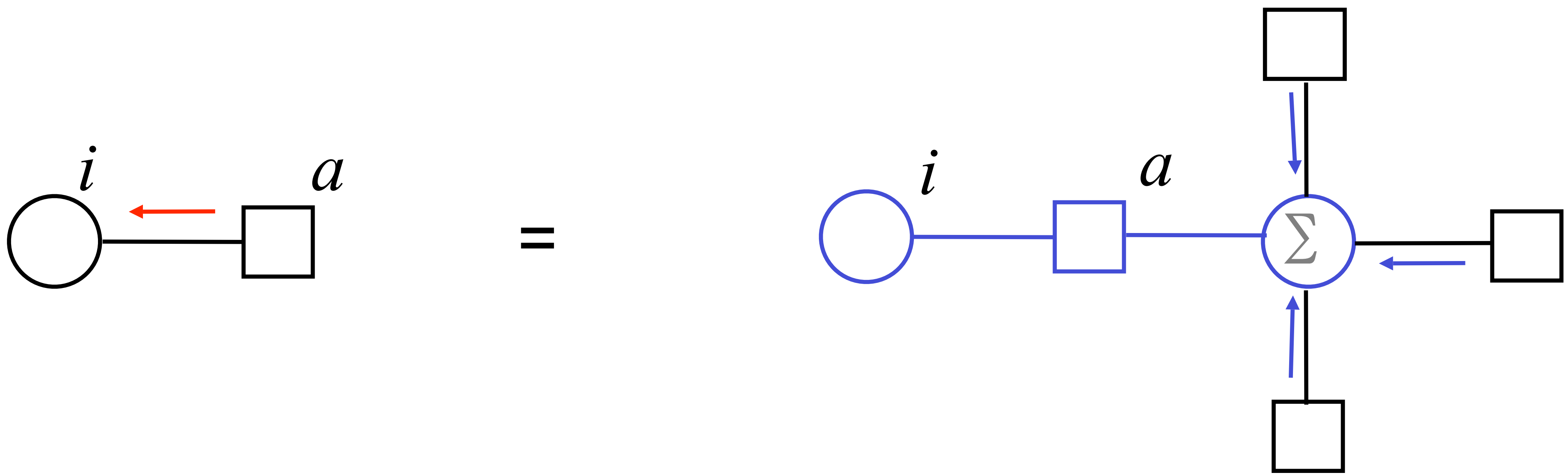
$$b_a(X_a) \propto f_a(X_a) \prod_{i \in N(a)} \prod_{b \in N(i) \setminus a} m_{b \to i}(x_i)$$

The "belief" is the BP approximation of the marginal probability.

BP Message-update Rules

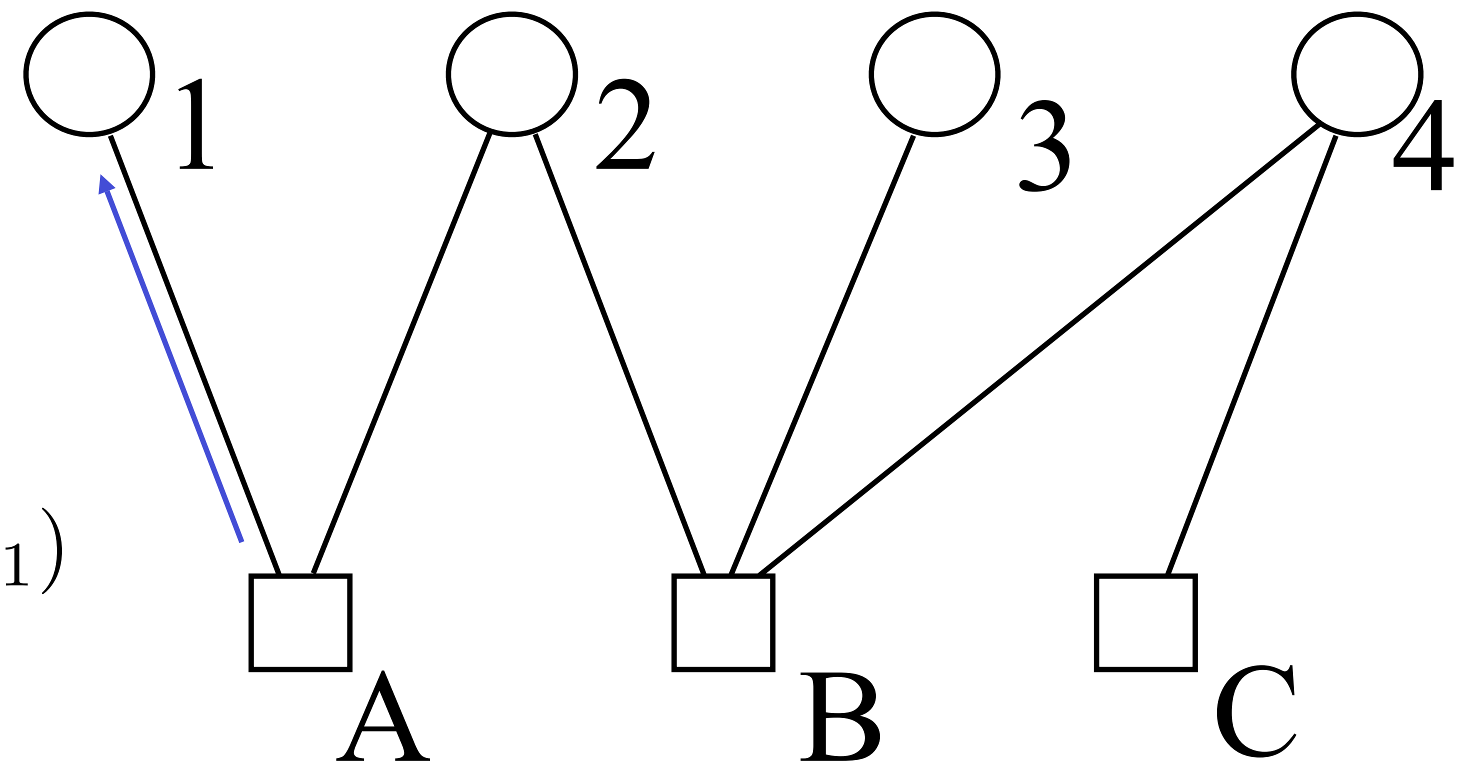
Using $b_i(x_i) = \sum_{X_a \setminus x_i} b_a(X_a)$, we get the "Sum-Product" message-update rules:

$$m_{a \to i}(x_i) = \sum_{X_a \setminus x_i} f_a(X_a) \prod_{j \in N(a) \setminus i} \prod_{b \in N(j) \setminus a} m_{b \to j}(x_j)$$

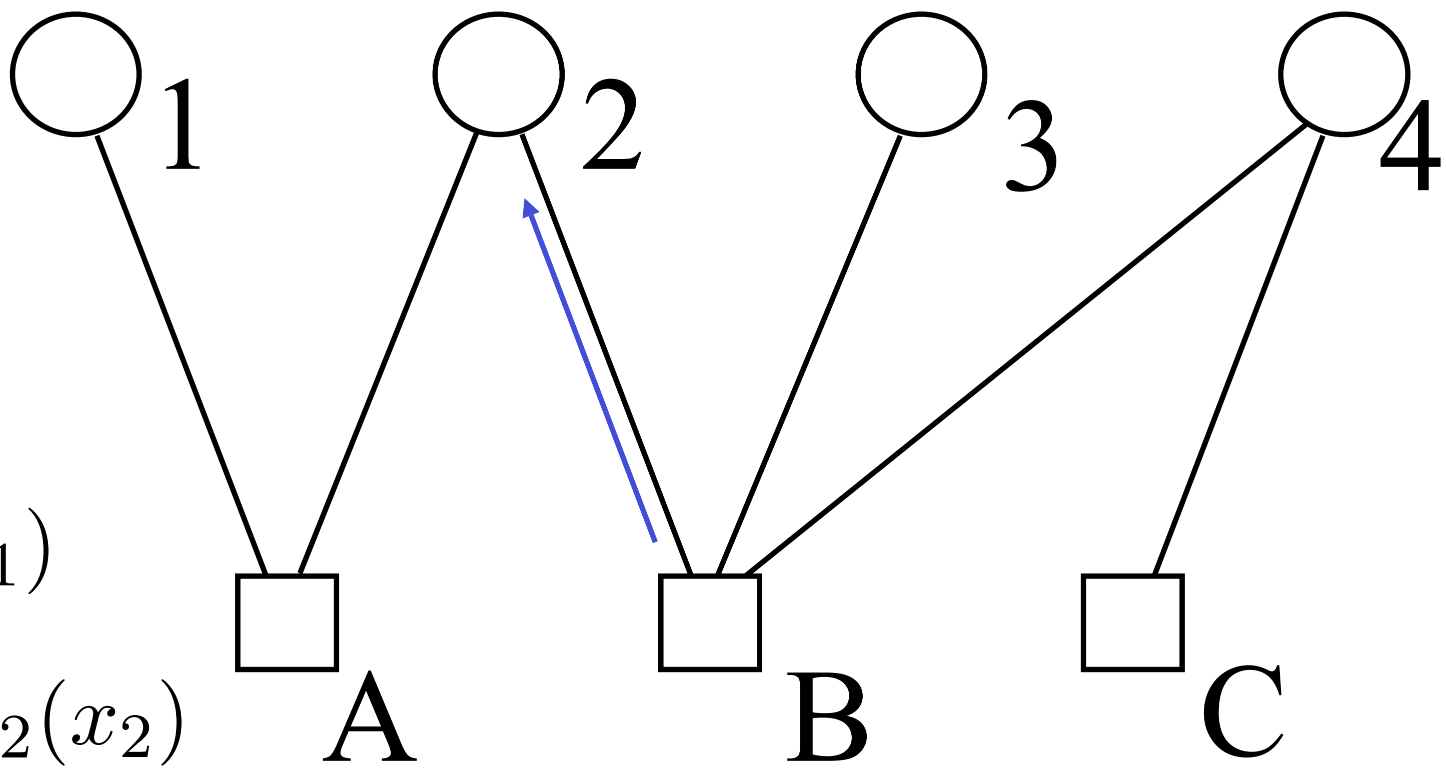


BP Is Exact for Trees

$b_1(x_1) \propto m_{A \to 1}(x_1)$



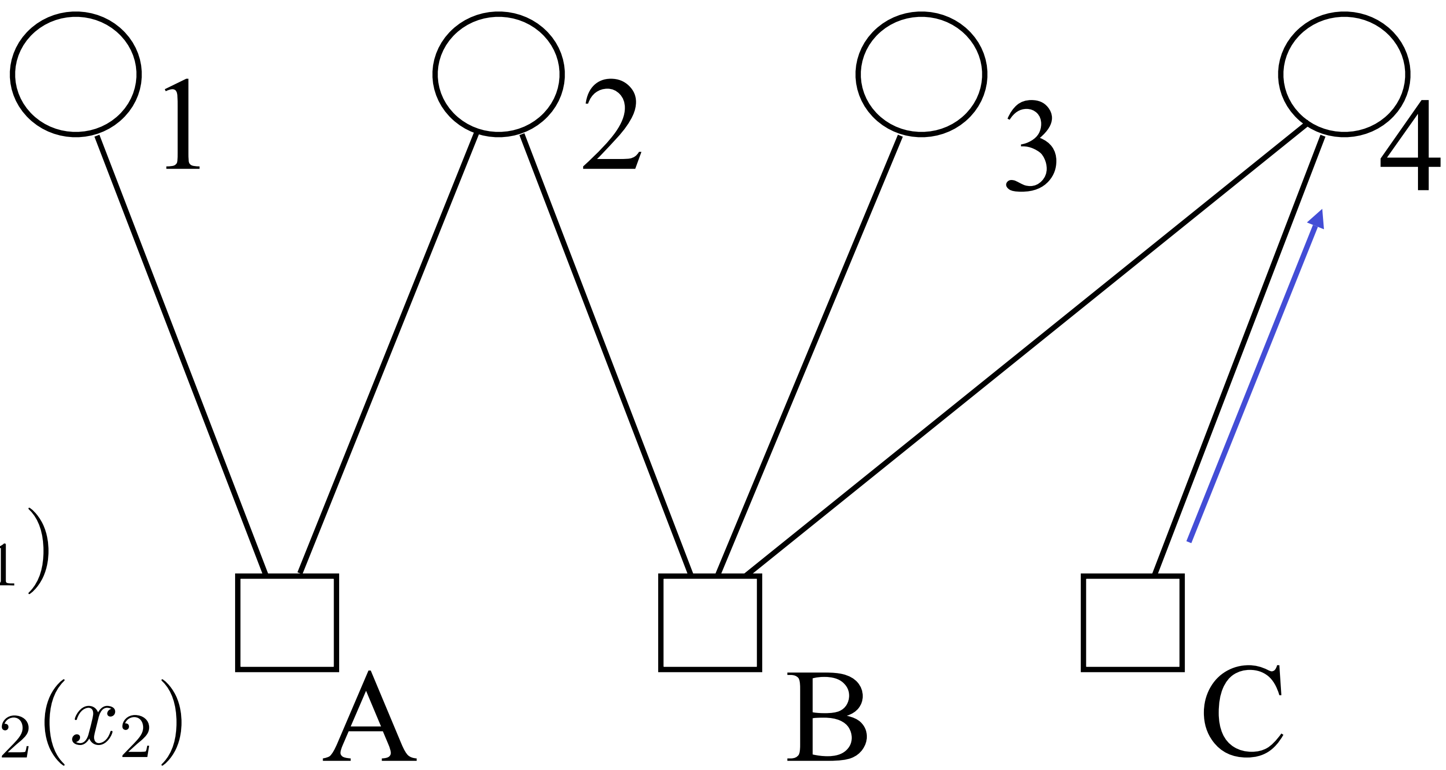
BP Is Exact for Trees



$$b_1(x_1) \propto m_{A\to 1}(x_1)$$

$$\propto \sum_{x_2} f_A(x_1, x_2) m_{B\to 2}(x_2)$$

BP Is Exact for Trees

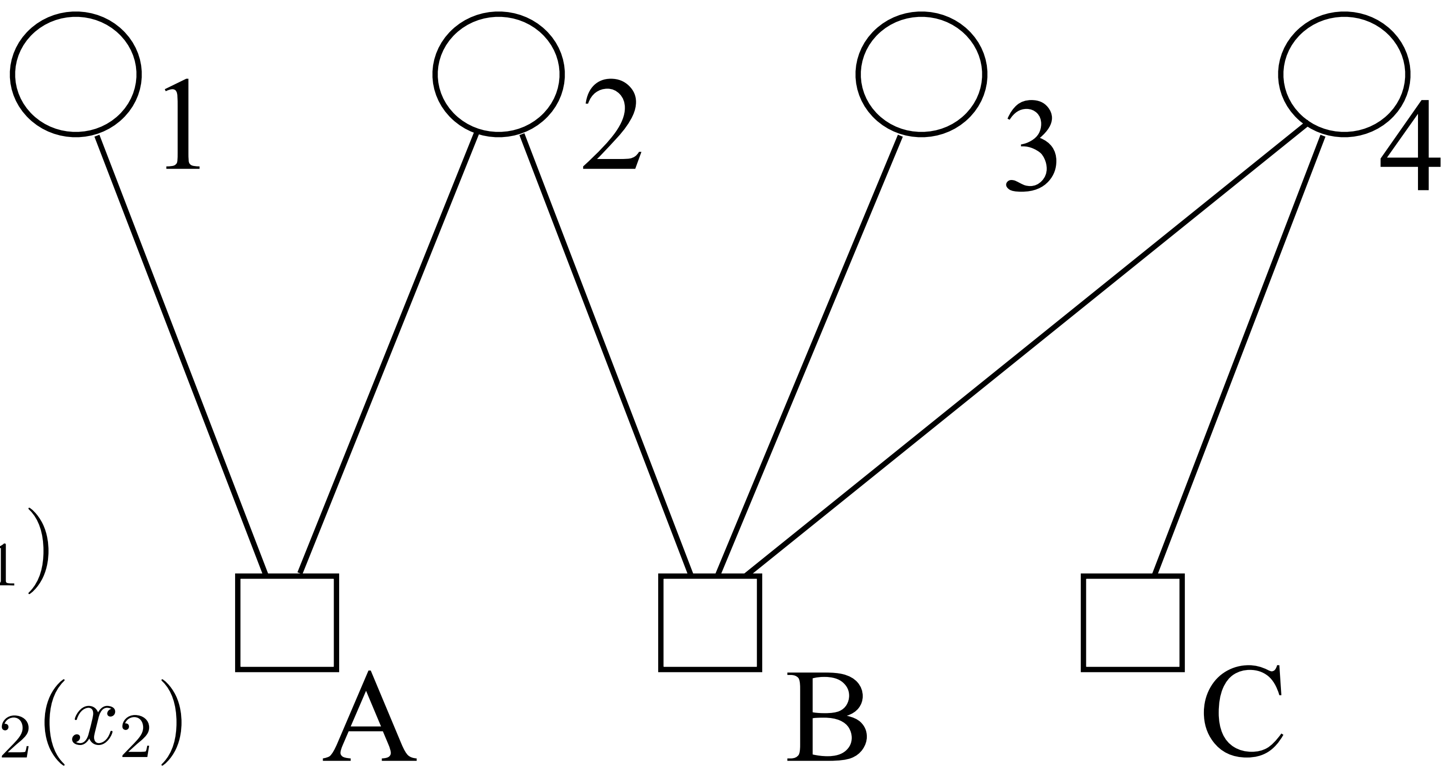


$$b_1(x_1) \propto m_{A\to 1}(x_1)$$

$$\propto \sum_{x_2} f_A(x_1, x_2) m_{B\to 2}(x_2)$$

$$\propto \sum_{x_2, x_3, x_4} f_A(x_1, x_2) f_B(x_2, x_3, x_4) m_{C\to 4}(x_4)$$

BP Is Exact for Trees



$$b_1(x_1) \propto m_{A\to1}(x_1)$$

$$\propto \sum_{x_2} f_A(x_1,x_2) m_{B\to2}(x_2)$$

$$\propto \sum_{x_2,x_3,x_4} f_A(x_1,x_2) f_B(x_2,x_3,x_4) m_{C\to4}(x_4)$$

$$\propto \sum_{x_2,x_3,x_4} f_A(x_1,x_2) f_B(x_2,x_3,x_4) f_C(x_4)$$

Max-Product BP

Max-Product (a.k.a. min-sum) message-update rules provably give the most probable configuration for graphs without cycles. This algorithm is also known as the Viterbi algorithm.

Outline

- Factor Graphs
- Message Passing Algorithms
- Free Energy Approximations

Variational Free Energies

$$F = U - TS = \sum_s p_s E_s + T \sum_s p_s \log p_s$$

Variational Free Energies

$$F = U - TS = \sum_s p_s E_s + T \sum_s p_s \log p_s$$

$$G(b_s) = \sum_s b_s E_s + T \sum_s b_s \log b_s$$

Variational Free Energies

$$F = U - TS = \sum_s p_s E_s + T \sum_s p_s \log p_s$$

$$G(b_s) = \sum_s b_s E_s + T \sum_s b_s \log b_s$$

Optimizing *G(b*$_s$*)*, with a normalized *b*$_s$, we obtain Boltzmann's Law:

$$b_s = \frac{1}{Z} \exp\left(-E_s/T\right)$$

Variational Derivation of Mean-Field Theory

Choose the b_s from a limited and tractable sub-space:

$$b_s = \prod_i b_i(x_i)$$

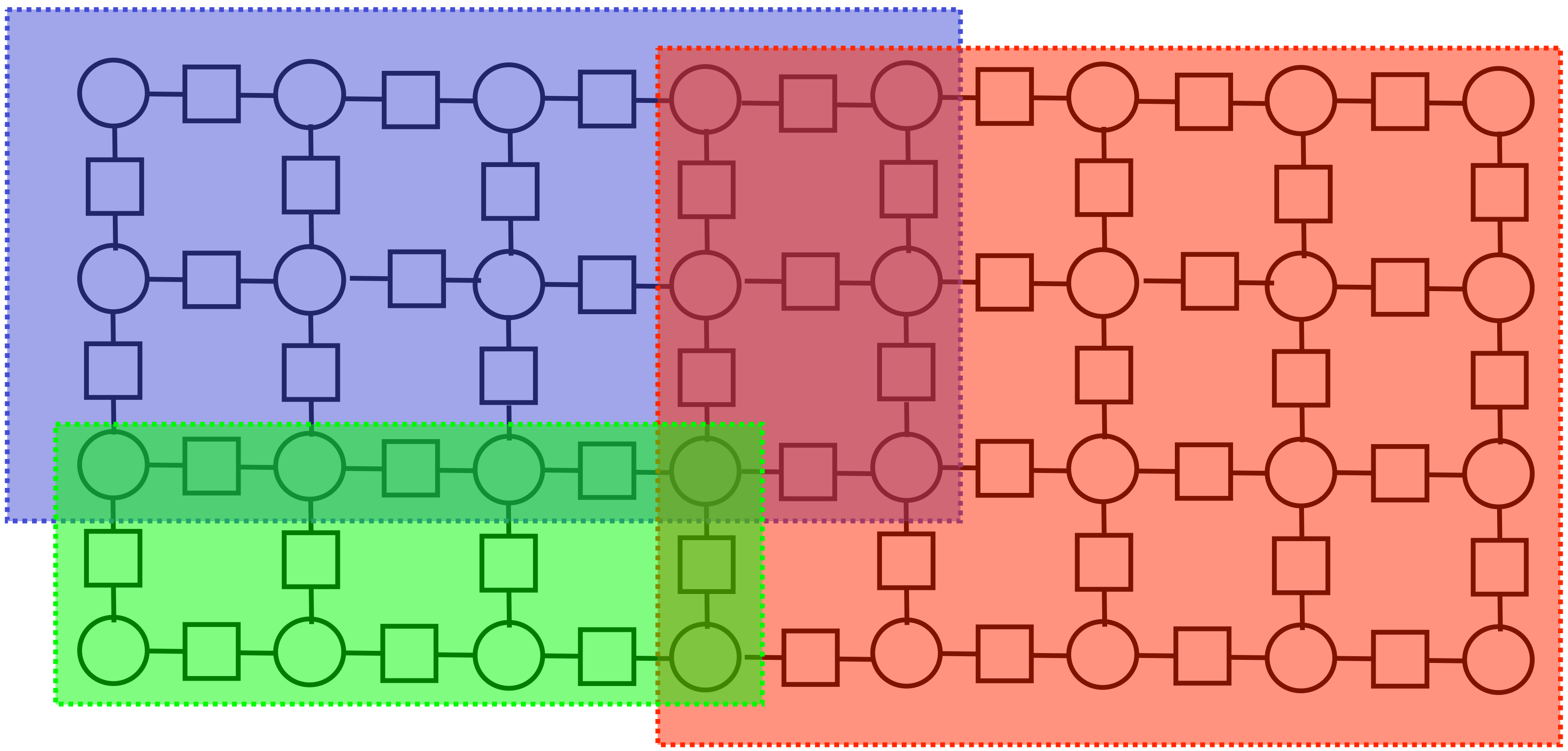
Minimizing the free energy over beliefs from this limited sub-space will give a rigorous upper-bound on the true free energy.

Region-based Approximations to the Variational Free Energy

(Kikuchi, 1951)

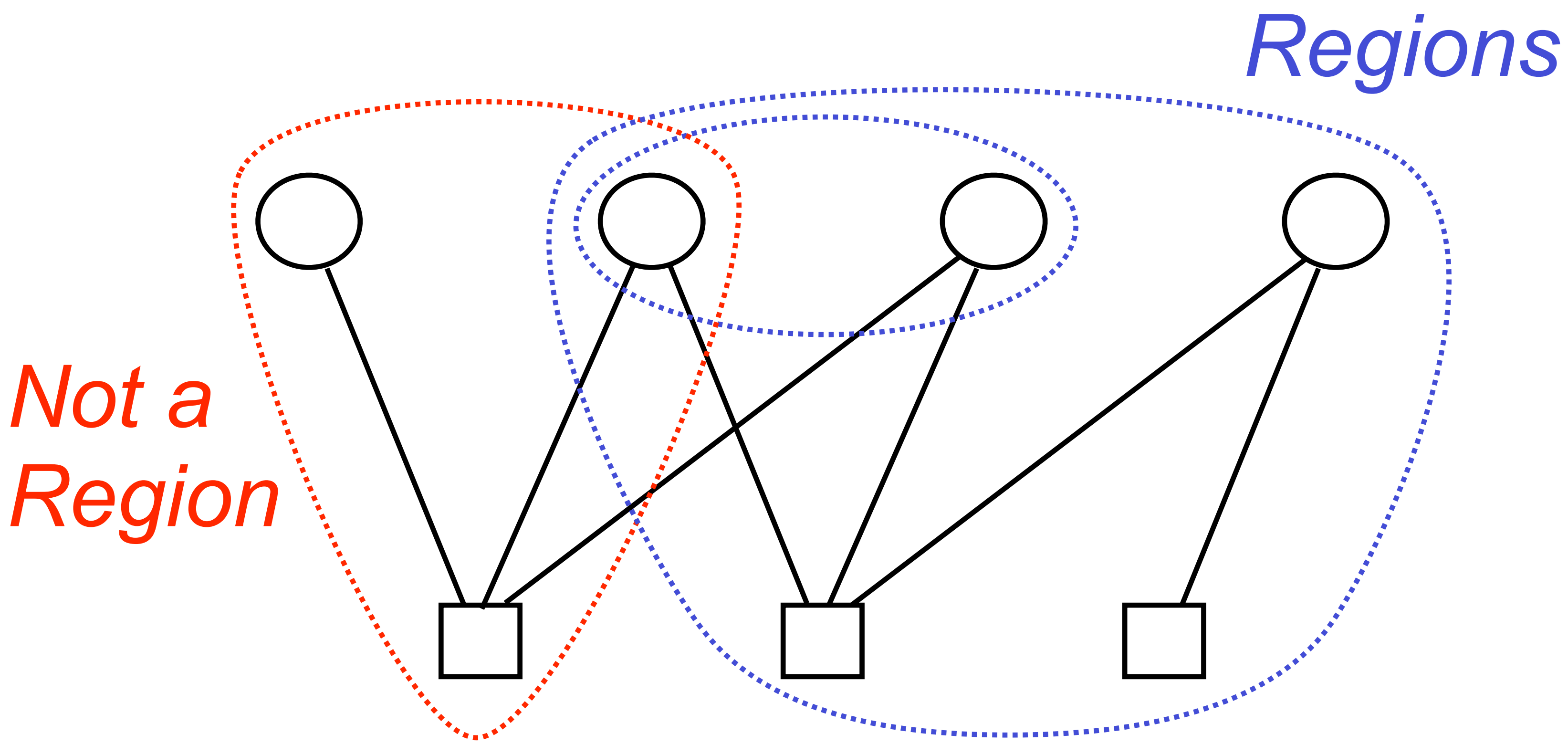
Exact: $G[b(X)]$ *(intractable)*

Regions: $G[\{b_r(X_r)\}]$



Defining a "Region"

A *region* r is a set of variable nodes V_r and factor nodes F_r such that if a factor node a belongs to F_r, all variable nodes neighboring a must belong to V_r. A region free energy can be naturally defined. The overall free energy will be the sum of the region free energies, weighted by their counting numbers.

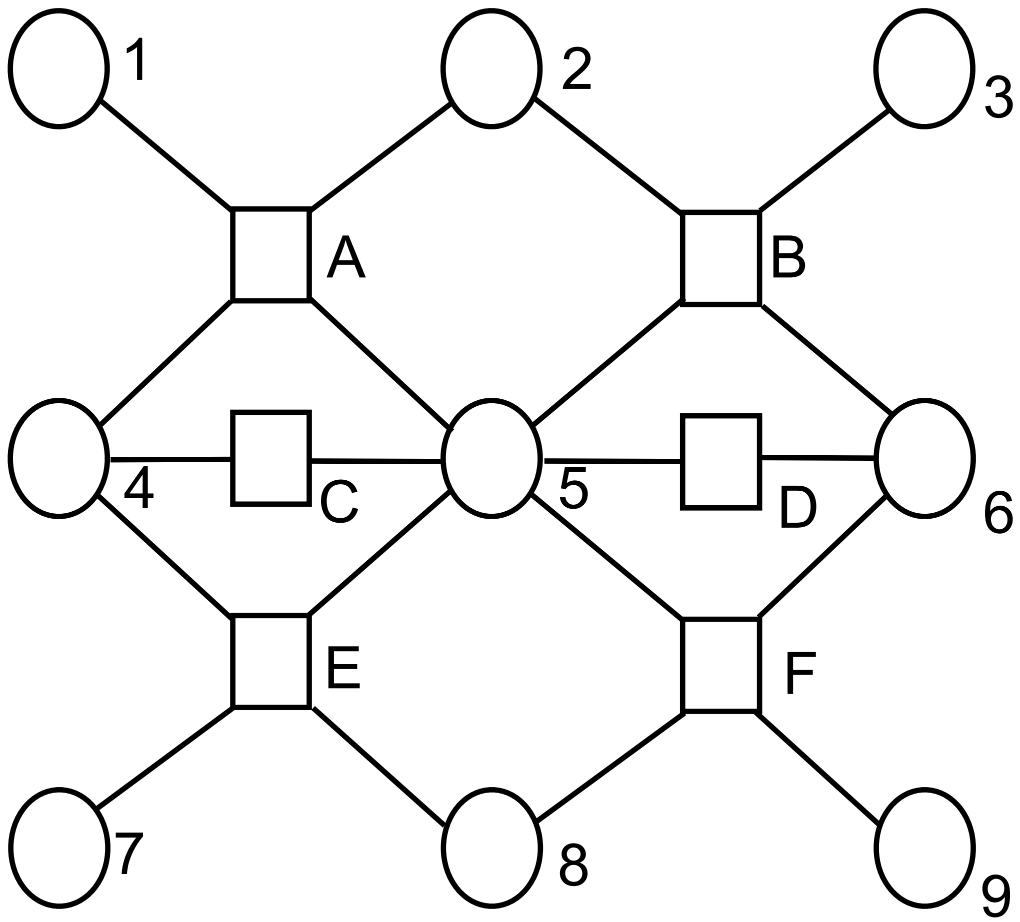


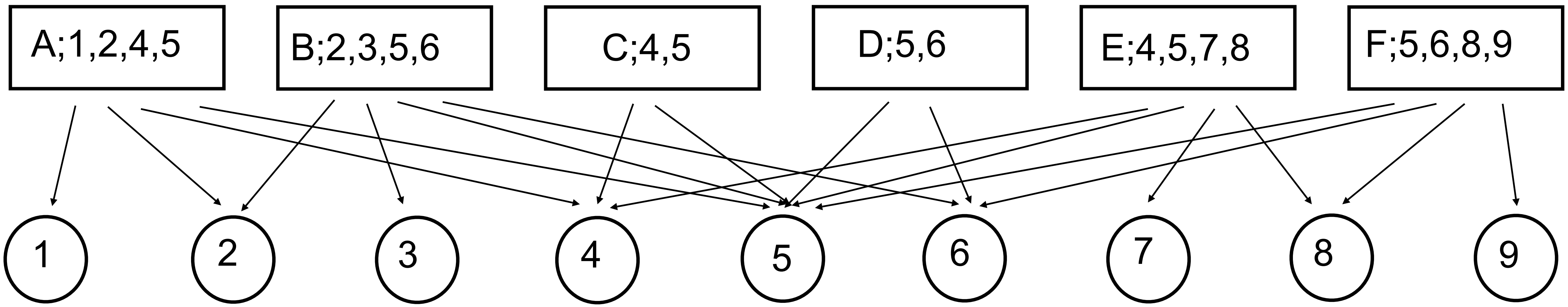
Bethe Approximation

Two sets of regions:

Large regions containing a single factor node *a* and all attached variable nodes. $c_r = 1$

Small regions containing a single variable node *i*. $c_r = 1 - d_i$





Bethe Free Energy

$$G_{Bethe} = \sum_a \sum_{X_a} b_a(X_a) \log \left(\frac{b_a(X_a)}{f_a(X_a)} \right) + \sum_i (1 - d_i) \sum_{x_i} b_i \log b_i(x_i)$$

Minimizing the Bethe Free Energy, subject to the normalization and consistency constraints, gives a set of equations that is equivalent to the Belief Propagation message update rules!

Improving on BP

- Generalized Belief Propagation: Using larger regions, construct a more accurate region-based free energy, or equivalent message-passing algorithm. Particularly useful for factor graphs defined on a square lattice.
- Survey Propagation: Account for the possibility of many thermodynamic states by keeping a probability distribution over messages. Particularly useful for constraint-satisfaction problems.
- Loop Calculus: An expansion where the zeroth order term is BP, and the corrections can be systematically computed.
- Convexified Free Energies: approximations, with a unique global minimum which can be used to obtain rigorous *lower* bounds on the free energy.
- Expectation Propagation: an algorithm, with a different derivation but very close connections to Generalized Belief Propagation; it also can be used to derive message-passing algorithms for factor graphs with continuous variables.

Pointers to the Literature

- “Information, Physics, and Computation,” M. Mézard and A. Montanari, Oxford University Press, forthcoming--draft available online.

- “Factor Graphs and the Sum-Product Algorithm,” F.R. Kschischang, B.J. Frey, and H.-A. Loeliger, *IEEE Trans. Info. Theory,* 47:498-519, (2001).

- “Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms,” J.S. Yedidia, W.T. Freeman, and Y. Weiss, *IEEE Trans. Info. Theory,* 51:2282-2312, (2005).

- “A New Class of Upper Bounds on the Log Partition Function,” M.J. Wainwright, T. S. Jaakkola, and Alan S. Willsky, *IEEE Trans. Info Theory,* 51: 2313-2335 (2005).

- “Structured Region Graphs: Morphing EP into GBP,” M. Welling, T. Minka, and Y.W. Teh, *UAI* (2005).

- “Survey Propagation: an Algorithm for Satisifiability,” A. Braunstein, M. Mézard, and R. Zecchina, *Random Structures and Algorithms,* 27: 201-226 (2005).

- “Loop Series for Discrete Statistical Models on Graphs” M. Chertkov and V. Chernyak, cond-mat/0603189, JSTAT P06009 (2006).

- and references in these papers and more recent papers by all these authors. (Apologies to all those I’ve neglected.)