From Belief Propagation to Generalised Belief Propagation

Mahdi Jafari Siavoshani
Sharif University of Technology

Institute of Network Coding
September 2015

Outline

- Backgrounds on Graphical Models
- Backgrounds on Statistical Physics
- Region-Based Approximation
- A Special Case: Bethe Approximation and Recovering BP
- Region Graph Method and Generalized Belief Propagation (GBP)
- GBP for Estimating the Partition Function of the 2D Ising Model

Backgrounds on Graphical Models

Graphical Models Is All About Factorization

- Consider n random variables $X_1, \dots, X_n$ where $X_i \in \mathcal{X}_i$

$$p(x_1, \dots, x_n) = \prod_{a \in A} \psi_a(x_a)$$

Probabilistic notions such as conditional independence

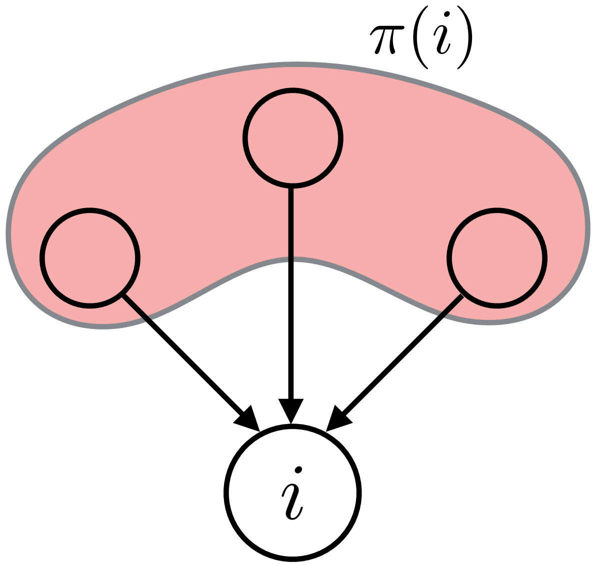
<==>

Graph-theoretic notions such as cliques and separation

- Generally two types of graphical models are common in practice
  - Bayesian Network (directed graphical models)
  - Markov Random Field (undirected graphical models)

Bayesian Network

- The probability distribution is factorized according to a directed acyclic graph



$$p(x_1, \ldots, x_n) = \prod_{i \in V} p_i(x_i | x_{\pi(i)})$$

$$p_i(x_i | x_{\pi(i)}) \geq 0$$

$$\int p_i(x_i | x_{\pi(i)}) = 1$$

- $p_i(x_i | x_{\pi(i)})$ is indeed a conditional probability distribution

Markov Random Field

- Let $G(V, E)$ be an undirected graph and $p(x_V) > 0$
  - Global Markov Property:

$$\forall W \subseteq V : \quad p(x_W | x_{V \setminus W}) = p(x_W | x_{\Delta W})$$

$$\Big\Updownarrow$$

Hammersley and Clifford Theorem

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

Markov Random Field

- Let $G(V, E)$ be an undirected graph and $p(x_V) > 0$

- Global Markov Property:

$$\forall W \subseteq V : \quad p(x_W | x_{V \setminus W}) = p(x_W | x_{\Delta W})$$

$$\Updownarrow$$

Hammersley and Clifford Theorem

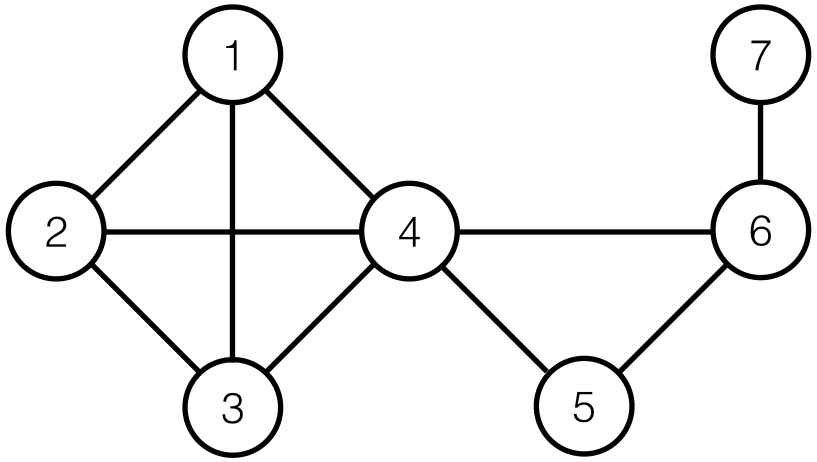
$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

Normalization constant called partition function

Usually the set of maximal cliques

Markov Random Field

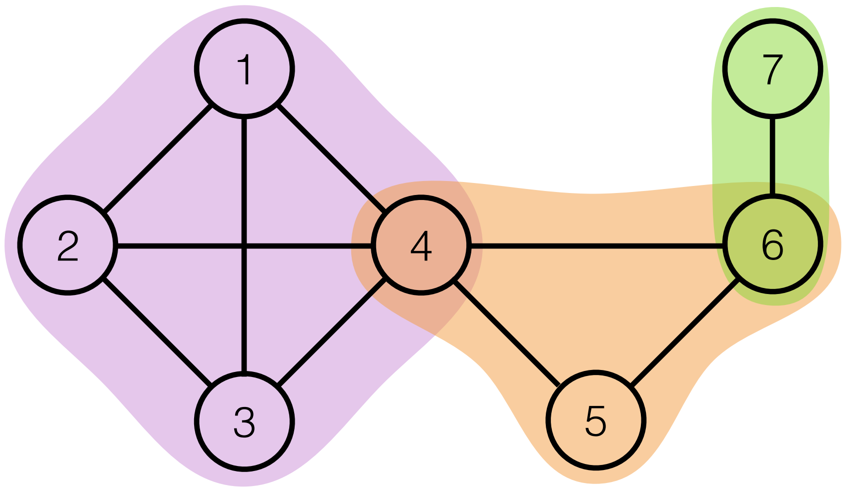
- Example:



$$p(x_1, \ldots, x_7) = \frac{1}{Z}\psi_{1234}(x_1, \ldots, x_4)\psi_{456}(x_4, x_5, x_6)\psi_{67}(x_6, x_7)$$

Markov Random Field

- Example:



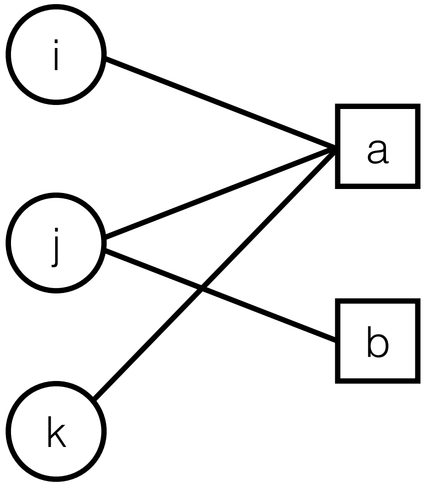
$$p(x_1, \ldots, x_7) = \frac{1}{Z}\psi_{1234}(x_1, \ldots, x_4)\psi_{456}(x_4, x_5, x_6)\psi_{67}(x_6, x_7)$$

Factor Graph

- Let $V = \{1, \ldots, n\}$ and A indexes the factors
  => A factor graph is a bipartite graph $G = (V, A, E)$

$$p(\boldsymbol{x}) \triangleq p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{a \in A} \psi_a(\boldsymbol{x}_a)$$

$$Z = \sum_{\boldsymbol{x}} \prod_{a \in A} \psi_a(\boldsymbol{x}_a)$$

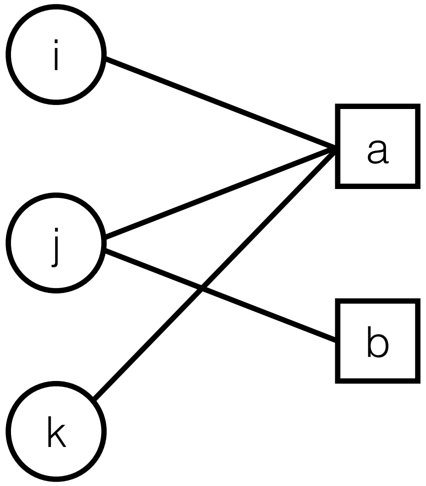


Factor Graph

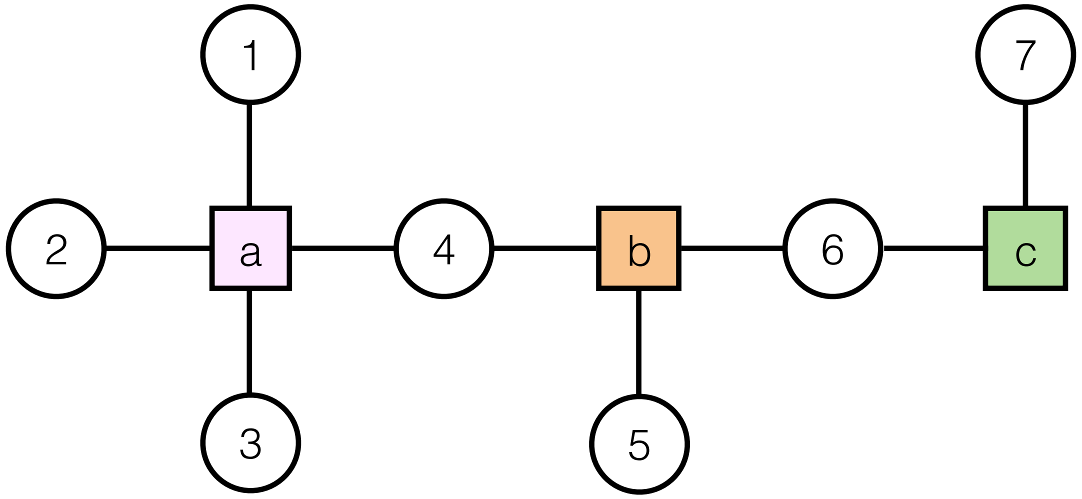
- Let $V = \{1, \ldots, n\}$ and A indexes the factors
  => A factor graph is a bipartite graph $G = (V, A, E)$

$$p(\boldsymbol{x}) \triangleq p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{a \in A} \psi_a(\boldsymbol{x}_a)$$

$$Z = \sum_{\boldsymbol{x}} \prod_{a \in A} \psi_a(\boldsymbol{x}_a)$$



- Example: $V = \{1, \ldots, 7\}$ and $A = \{a, b, c\}$



$$p(\boldsymbol{x}) = \frac{1}{Z} \psi_a(\boldsymbol{x}_a) \psi_b(\boldsymbol{x}_b) \psi_c(\boldsymbol{x}_c)$$

Two Important Problems!

- Computing the marginal distribution $p(\boldsymbol{x}_W)$ over a particular subset $W \subset V$ of nodes

$$p(\boldsymbol{x}_W) = \sum_{\boldsymbol{x} \setminus \boldsymbol{x}_W} p(\boldsymbol{x})$$

- Computing a mode of the density

$$\arg\max_{\boldsymbol{x} \in \mathcal{X}^n} p(\boldsymbol{x})$$

Two Important Problems!

- Computing the marginal distribution $p(\boldsymbol{x}_W)$ over a particular subset $W \subset V$ of nodes

$$p(\boldsymbol{x}_W) = \sum_{\boldsymbol{x} \setminus \boldsymbol{x}_W} p(\boldsymbol{x})$$

- Computing a mode of the density

$$\arg\max_{\boldsymbol{x} \in \mathcal{X}^n} p(\boldsymbol{x})$$

In general, these problems are hard!

- Example: Consider binary random variables $X_0, \ldots, X_{100}$. To compute $p(x_0)$ we need to sum over an exponential number of terms:

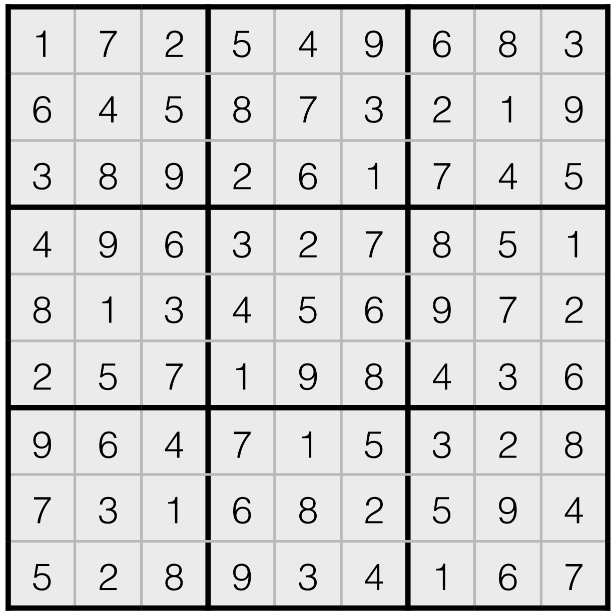
$$p(x_0) = \sum_{x_1, \ldots, x_{100} \in \{0,1\}} p(x_0, x_1, \ldots, x_{100})$$

Partition Function

- The partition function Z of a graphical model encodes important information about the underlying distribution
  - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
  - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system

Partition Function

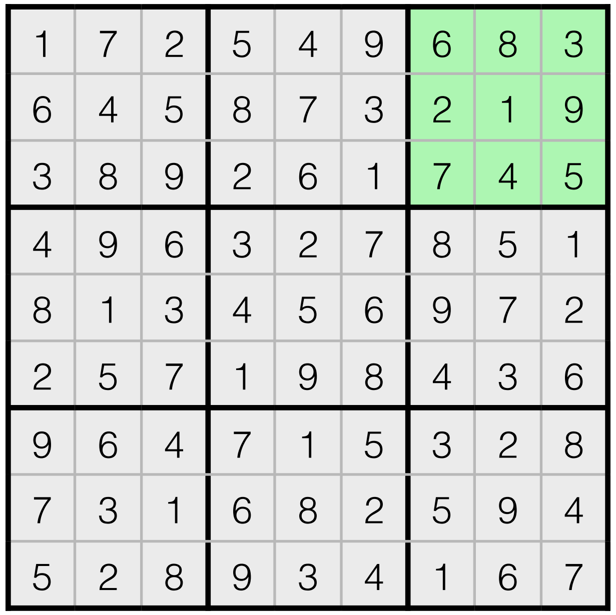
- The partition function Z of a graphical model encodes important information about the underlying distribution
  - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
  - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system



1	7	2	5	4	9	6	8	3
6	4	5	8	7	3	2	1	9
3	8	9	2	6	1	7	4	5
4	9	6	3	2	7	8	5	1
8	1	3	4	5	6	9	7	2
2	5	7	1	9	8	4	3	6
9	6	4	7	1	5	3	2	8
7	3	1	6	8	2	5	9	4
5	2	8	9	3	4	1	6	7

Partition Function

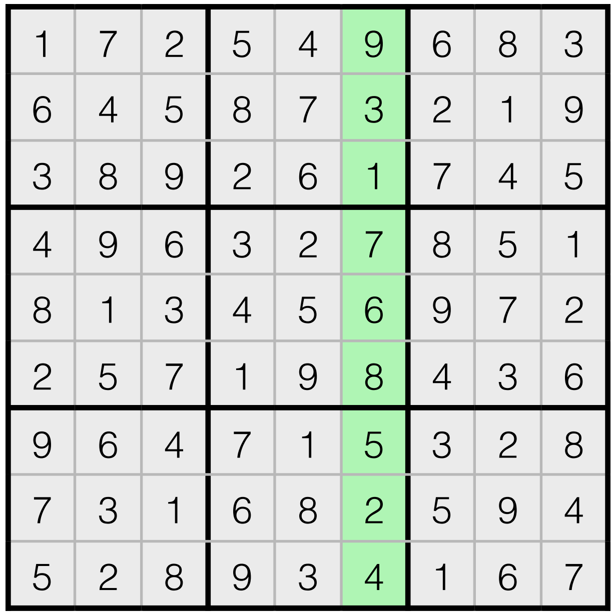
- The partition function Z of a graphical model encodes important information about the underlying distribution
  - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
  - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system



1	7	2	5	4	9	6	8	3
6	4	5	8	7	3	2	1	9
3	8	9	2	6	1	7	4	5
4	9	6	3	2	7	8	5	1
8	1	3	4	5	6	9	7	2
2	5	7	1	9	8	4	3	6
9	6	4	7	1	5	3	2	8
7	3	1	6	8	2	5	9	4
5	2	8	9	3	4	1	6	7

Partition Function

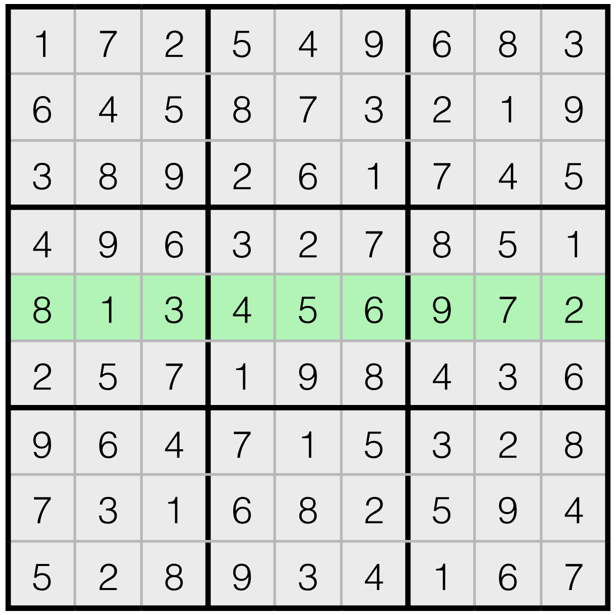
- The partition function Z of a graphical model encodes important information about the underlying distribution
  - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
  - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system



1	7	2	5	4	9	6	8	3
6	4	5	8	7	3	2	1	9
3	8	9	2	6	1	7	4	5
4	9	6	3	2	7	8	5	1
8	1	3	4	5	6	9	7	2
2	5	7	1	9	8	4	3	6
9	6	4	7	1	5	3	2	8
7	3	1	6	8	2	5	9	4
5	2	8	9	3	4	1	6	7

Partition Function

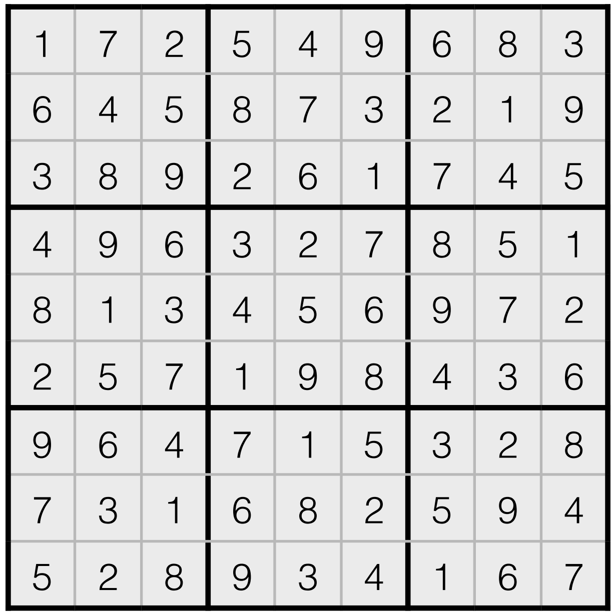
- The partition function Z of a graphical model encodes important information about the underlying distribution
    - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
    - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system



1	7	2	5	4	9	6	8	3
6	4	5	8	7	3	2	1	9
3	8	9	2	6	1	7	4	5
4	9	6	3	2	7	8	5	1
8	1	3	4	5	6	9	7	2
2	5	7	1	9	8	4	3	6
9	6	4	7	1	5	3	2	8
7	3	1	6	8	2	5	9	4
5	2	8	9	3	4	1	6	7

Partition Function

- The partition function Z of a graphical model encodes important information about the underlying distribution
  - Z is an important quantity for physicist => from Z we can compute experimentally measurable quantities
  - If all ψ_a are hard constraints => Z counts the number of valid configuration in the system



1	7	2	5	4	9	6	8	3
6	4	5	8	7	3	2	1	9
3	8	9	2	6	1	7	4	5
4	9	6	3	2	7	8	5	1
8	1	3	4	5	6	9	7	2
2	5	7	1	9	8	4	3	6
9	6	4	7	1	5	3	2	8
7	3	1	6	8	2	5	9	4
5	2	8	9	3	4	1	6	7

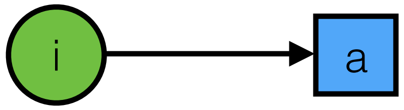
$Z =$ number of valid Sudoku configurations

Belief Propagation (BP)

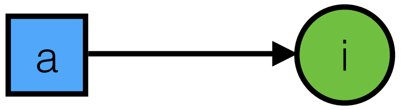
(Sum-Product Algorithm)

- Messages are exchanged between variable nodes and factor nodes of a factor graph

$n_{i \to a}(x_i)$



$m_{a \to i}(x_i)$

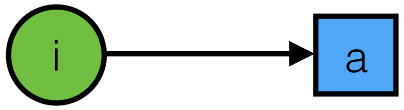


Belief Propagation (BP)

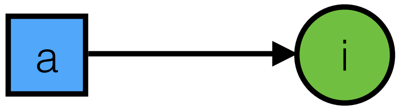
(Sum-Product Algorithm)

- Messages are exchanged between variable nodes and factor nodes of a factor graph

$n_{i\rightarrow a}(x_i)$



$m_{a\rightarrow i}(x_i)$

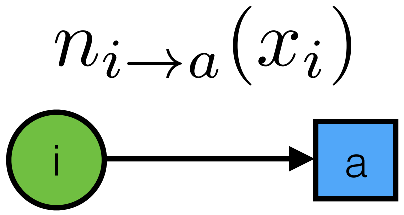


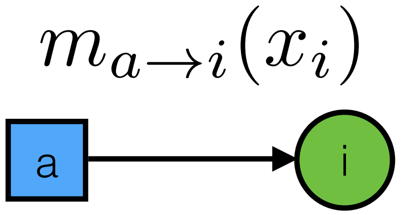
- Message update rules:

Belief Propagation (BP)

(Sum-Product Algorithm)

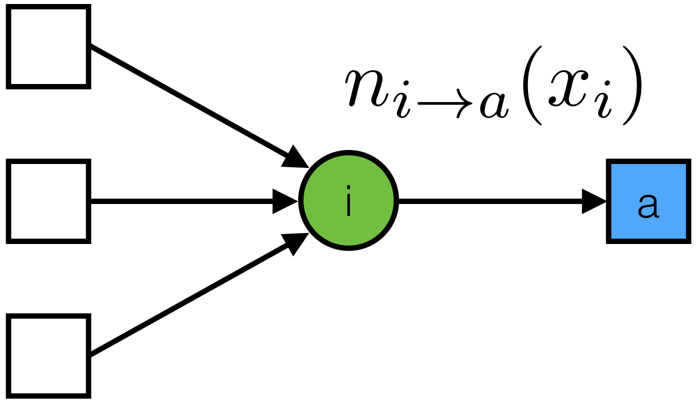
- Messages are exchanged between variable nodes and factor nodes of a factor graph





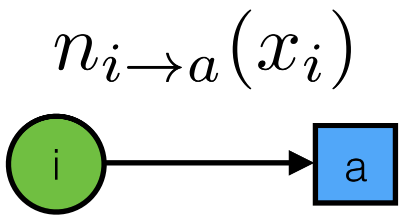
- Message update rules:

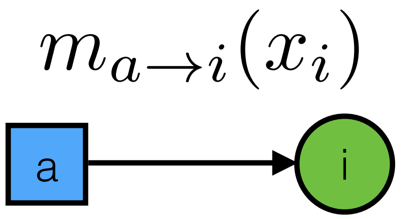
$$n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$$



Belief Propagation (BP)
(Sum-Product Algorithm)

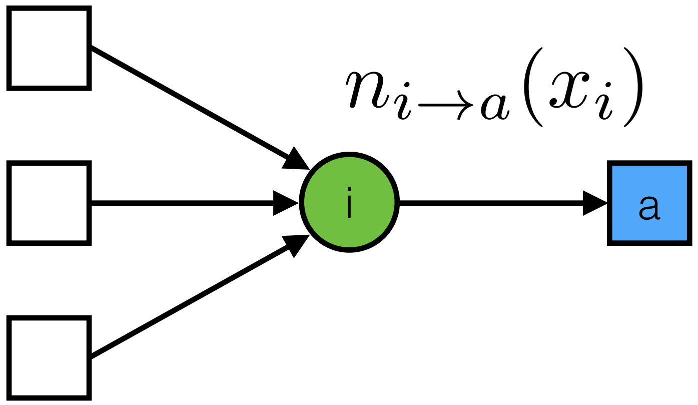
- Messages are exchanged between variable nodes and factor nodes of a factor graph

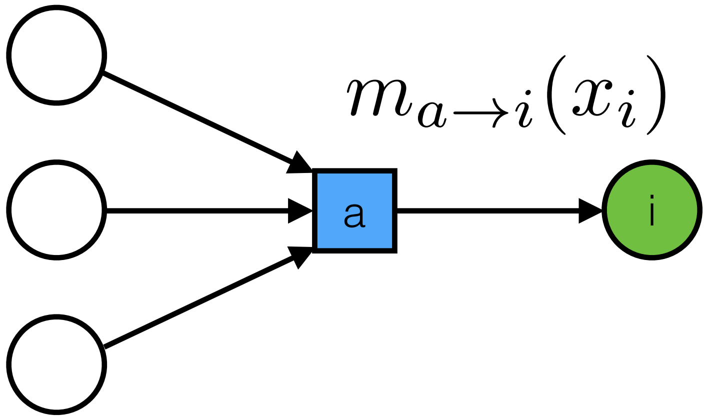




- Message update rules:

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



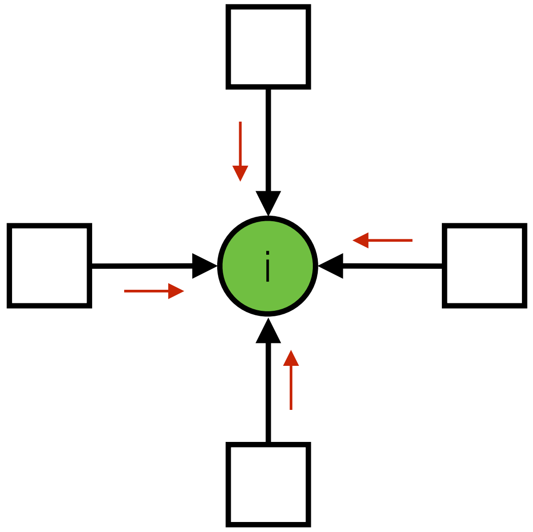


Belief Propagation (BP)

- How to compute the marginals?

Belief Propagation (BP)

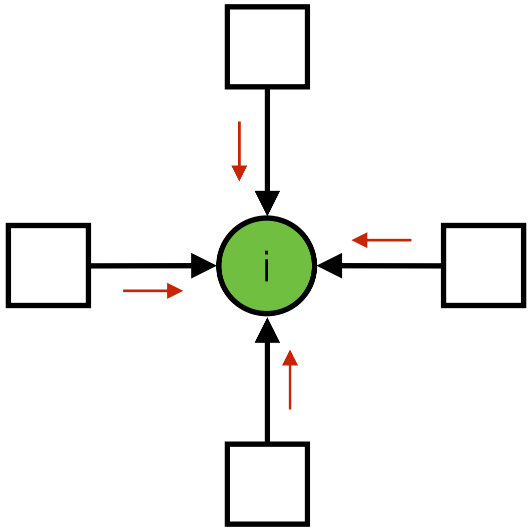
- How to compute the marginals?



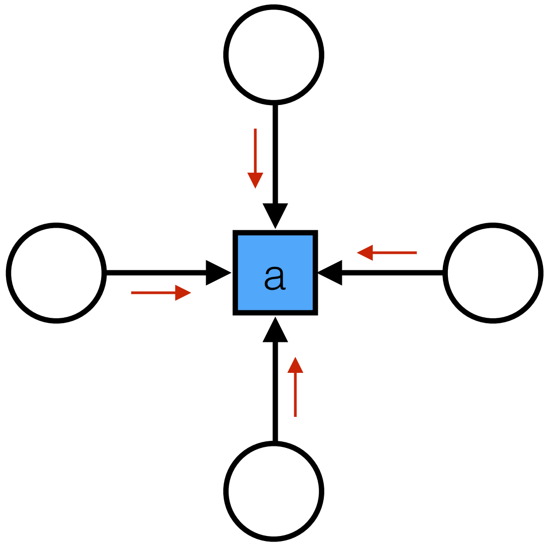
$$b_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$$

Belief Propagation (BP)

- How to compute the marginals?



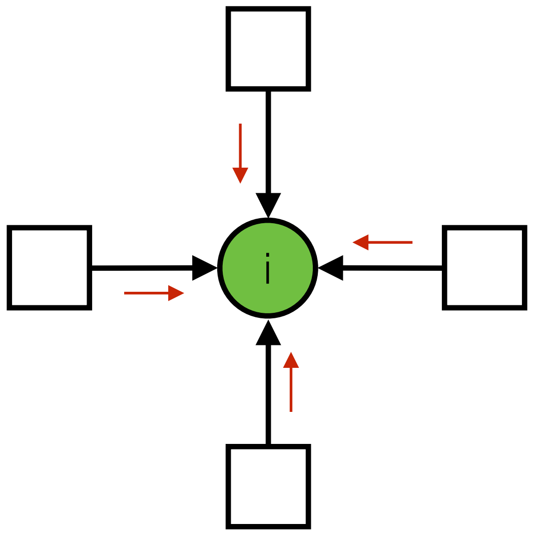
$$b_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$$



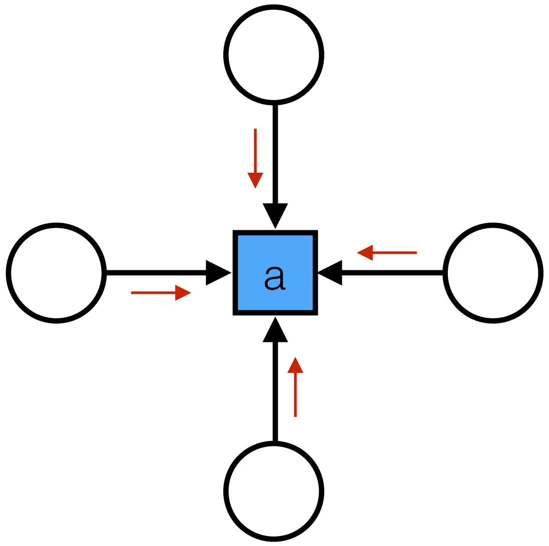
$$b_a(\boldsymbol{x}_a) \propto f_a(\boldsymbol{x}_a) \prod_{i \in N(a)} n_{i \to a}(x_i)$$

Belief Propagation (BP)

- How to compute the marginals?



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



$$b_a(\boldsymbol{x}_a) \propto f_a(\boldsymbol{x}_a) \prod_{i \in N(a)} n_{i \to a}(x_i)$$

BP is exact on trees, but only gives an approximation on graphs with cycles!

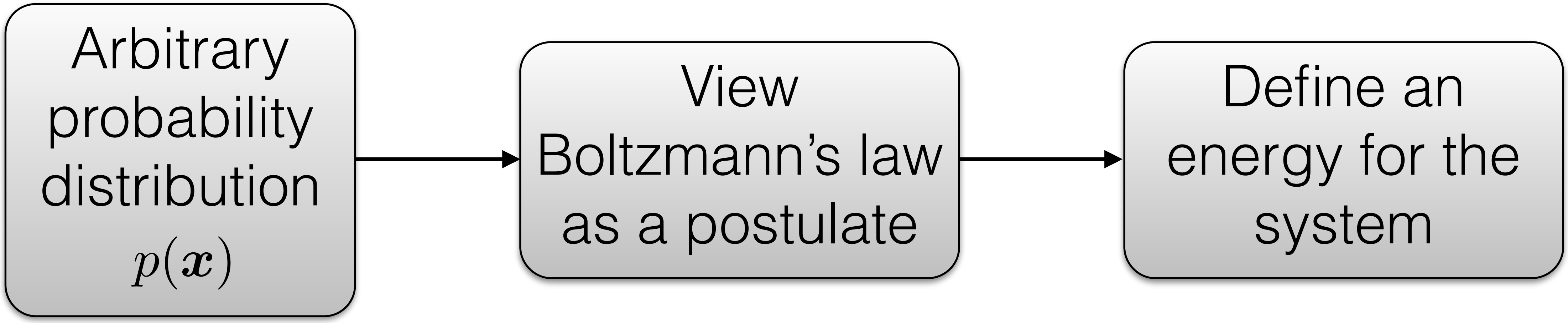
Backgrounds
on
Statistical Physics

Boltzmann Law

- A fundamental result of statistical mechanics is that, in thermal equilibrium, the probability of a state will be given by Boltzmann's distribution:

$$p(\boldsymbol{x}) = \frac{1}{Z(T)} e^{-E(\boldsymbol{x})/T}$$

Alternative point of view



Arbitrary probability distribution $p(\boldsymbol{x})$

View Boltzmann's law as a postulate

Define an energy for the system

Energy Assigned to a Factor Graph

- Consider factor graph $G = (V, A, E)$
  - For probability distribution

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_{a \in A} f_a(\boldsymbol{x}_a)$$

we can define energy of state $\boldsymbol{x}$ as

$$E(\boldsymbol{x}) = -\sum_{a \in A} \ln f_a(\boldsymbol{x}_a)$$

(Helmholtz) Free Energy

- Free energy of a system is defined as

$$F_H \triangleq U - H$$

- U is average energy:

$$U \triangleq \sum_{\boldsymbol{x}} p(\boldsymbol{x}) E(\boldsymbol{x})$$

- H is entropy:

$$H = -\sum_{\boldsymbol{x}} p(\boldsymbol{x}) \ln p(\boldsymbol{x})$$

- $p(\boldsymbol{x})$ is the actual probability distribution of the system
- Note that we have $F_H = -\ln Z$

Variational Approach

(Gibbs Free Energy)

- Instead of true probability distribution $p(\boldsymbol{x})$ consider some other distribution $b(\boldsymbol{x})$. Then define

$$F(b) \triangleq U(b) - H(b)$$

- where

$$U(b) \triangleq \sum_{\boldsymbol{x}} b(\boldsymbol{x}) E(\boldsymbol{x})$$

$$H(b) \triangleq \sum_{\boldsymbol{x}} b(\boldsymbol{x}) \ln b(\boldsymbol{x})$$

- We can show

$$F(b) = F_H + D(b \| p)$$

=> $F(b)$ takes its minimum at $b(\boldsymbol{x}) = p(\boldsymbol{x})$

Variational Approach

- Consider the following optimization problem

$$F_H = \left\{ \begin{array}{c} \min F(b) \\ \text{s.t. } \; b \text{ is a joint probability distribution over } \boldsymbol{x} \end{array} \right.$$

- This optimization problem provides an exact procedure for computing the partition function (in fact F_H) and recovering $p(\boldsymbol{x})$

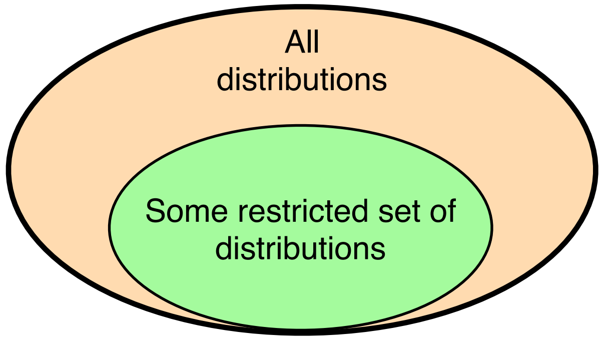
- Bad news: this problem is at least as hard as the original problem of partition function computation

As n becomes large, this method is intractable!

- Good news: we can use it to develop approximation methods!

A General Approach to Upper Bound F_H

- A more practical approach to upper bound F_H is to minimize $F(b)$ over a restricted class of probability distribution



- Example: mean-field approximation

$$b_{\mathrm{MF}} = \prod_{i \in V} b_i(x_i)$$

- We can extend this method by considering more complicated form for $b(\boldsymbol{x})$ that leads to a tractable distribution.
  => Example: structured mean-field approach

A General Approximation Approach

$$\begin{aligned} &\min_b F(b) \\ \text{s.t.} \quad &0 \leq b(\boldsymbol{x}) \leq 1, \quad \forall \boldsymbol{x} \\ &\sum_{\boldsymbol{x}} b(\boldsymbol{x}) = 1 \end{aligned}$$

A General Approximation Approach

$$\boxed{\min_{b} F(b)}$$

$$\text{s.t.} \quad 0 \leq b(\boldsymbol{x}) \leq 1, \quad \forall \boldsymbol{x}$$

$$\sum_{\boldsymbol{x}} b(\boldsymbol{x}) = 1$$

Approximate $F(b)$ by some simpler and more tractable function

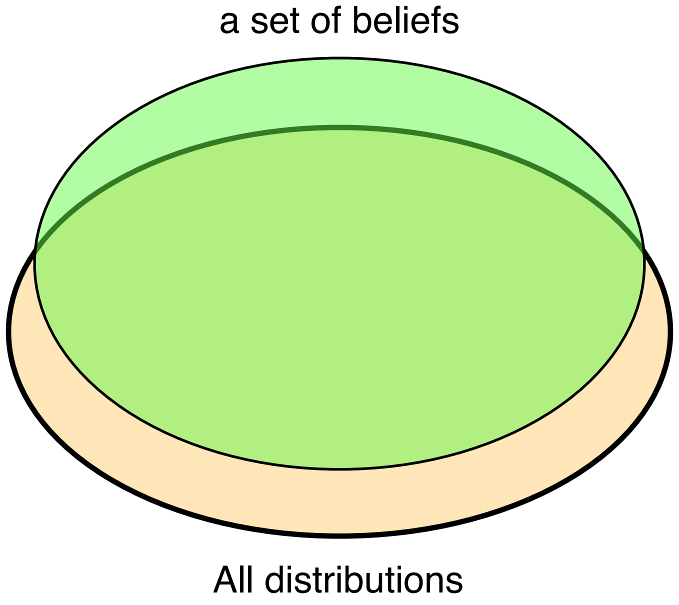
A General Approximation Approach

$$\boxed{\min_b F(b)}$$

$$\text{s.t.} \quad \boxed{\begin{aligned} &0 \le b(\boldsymbol{x}) \le 1, \quad \forall \boldsymbol{x} \\ &\sum_{\boldsymbol{x}} b(\boldsymbol{x}) = 1 \end{aligned}}$$

Approximate $F(b)$ by some simpler and more tractable function

Minimize $F(b)$ over a set of beliefs b which approximate the probability simplex over $\boldsymbol{x}$

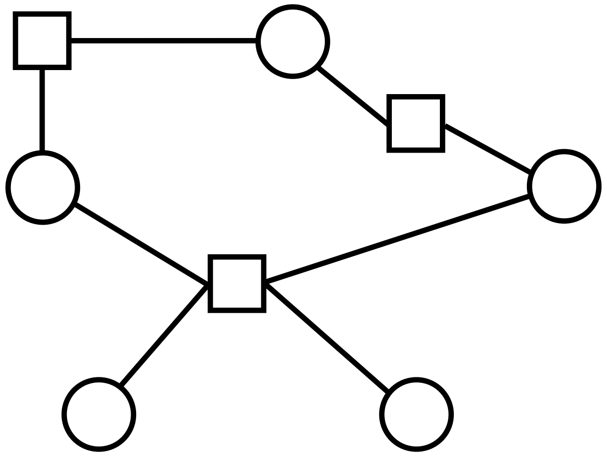


Region-Based Approximation

Region-Based Approximation

(Main Idea)

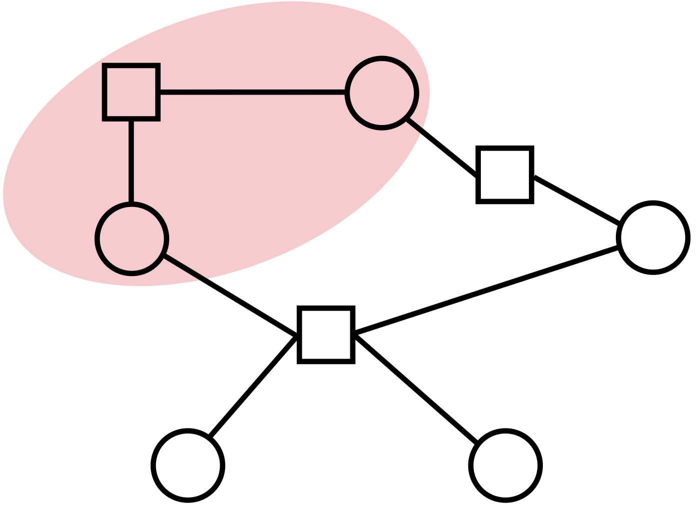
- Break the factor graph into regions



Region-Based Approximation

(Main Idea)

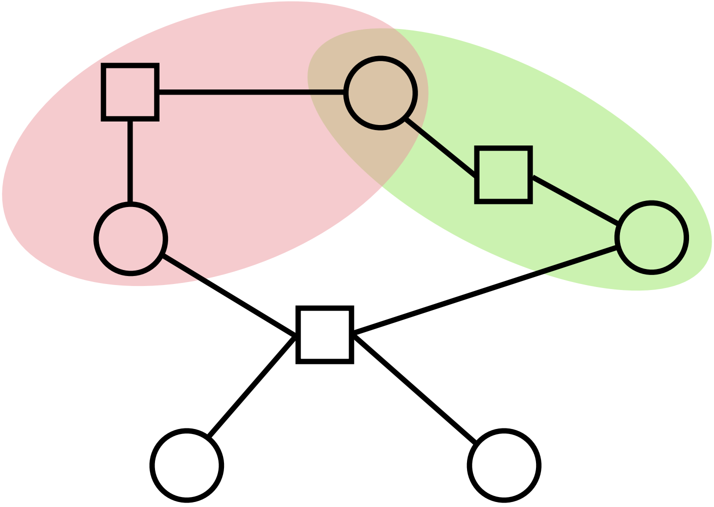
- Break the factor graph into regions



Region-Based Approximation

(Main Idea)

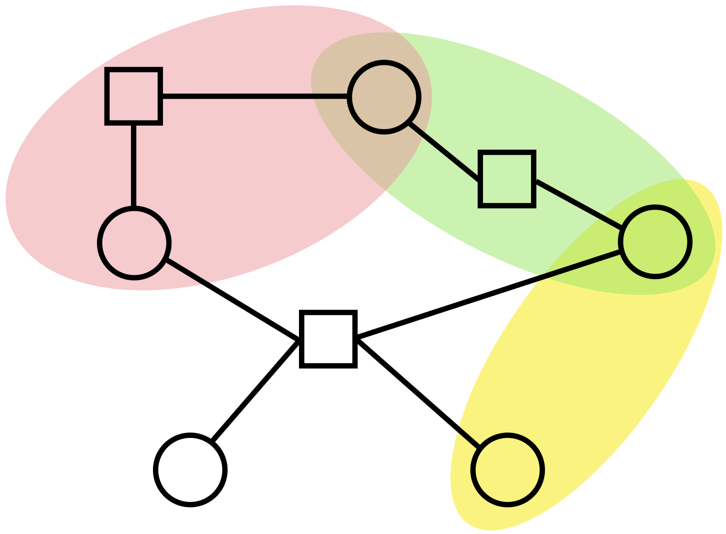
- Break the factor graph into regions



Region-Based Approximation

(Main Idea)

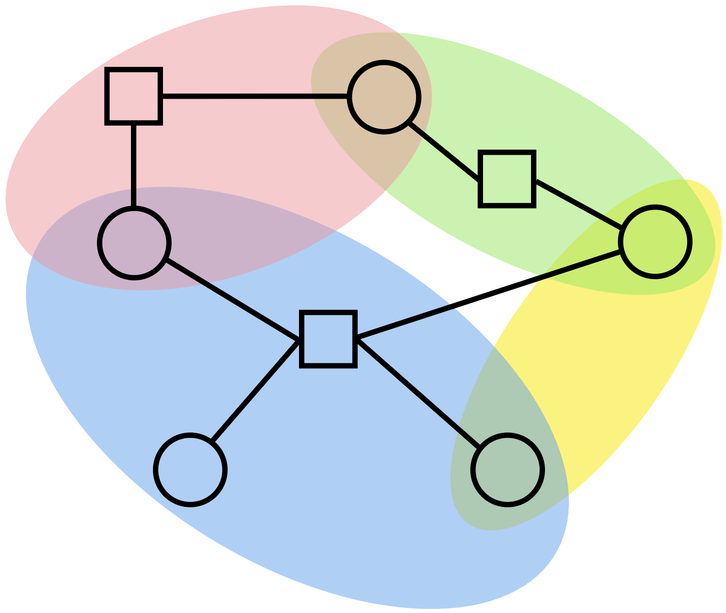
- Break the factor graph into regions



Region-Based Approximation

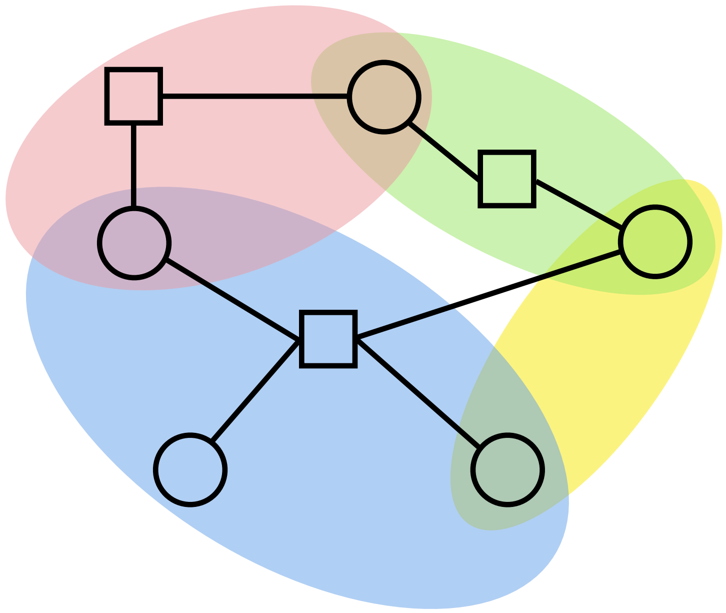
(Main Idea)

- Break the factor graph into regions



Region-Based Approximation
(Main Idea)

- Break the factor graph into regions



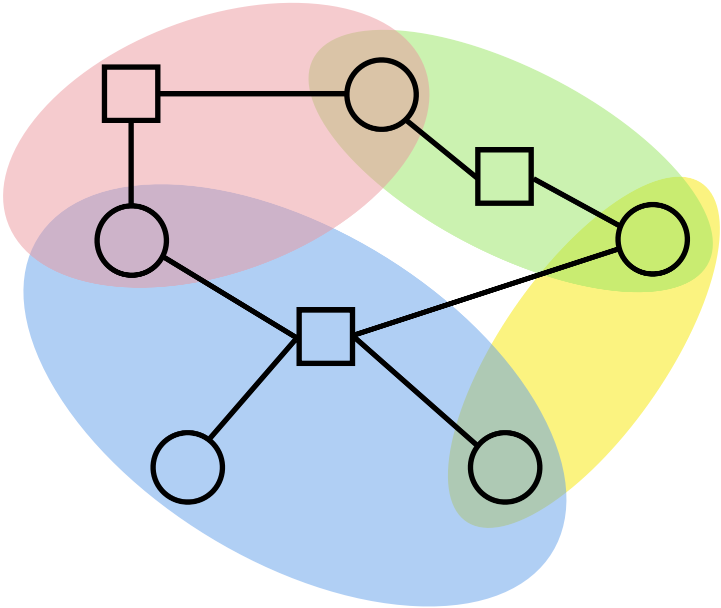
- Approximate the overall free energy as: the sum of the free energy of all the regions

$$F_{\mathcal{R}} \approx \sum_{R \in \mathcal{R}} F_R(b_R)$$

Region-Based Approximation

(Main Idea)

- Break the factor graph into regions



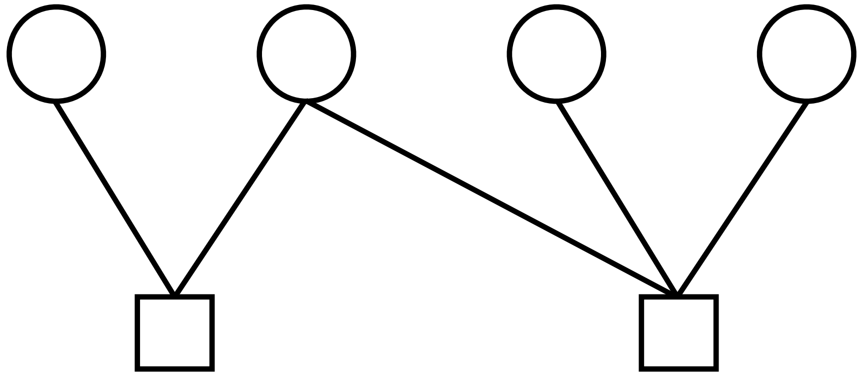
- Approximate the overall free energy as: the sum of the free energy of all the regions

$$F_{\mathcal{R}} \approx \sum_{R \in \mathcal{R}} F_R(b_R)$$

- Heuristic: to have a good approximation => Find good set of regions

Region-Based Approximation
(Definitions)

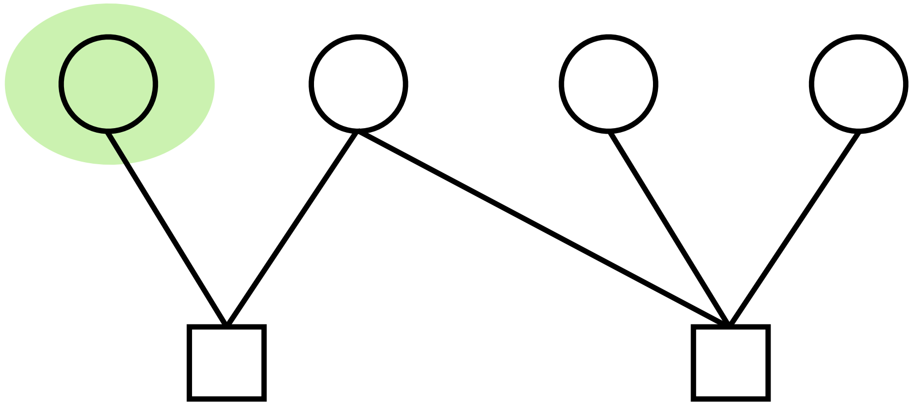
- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



Region-Based Approximation

(Definitions)

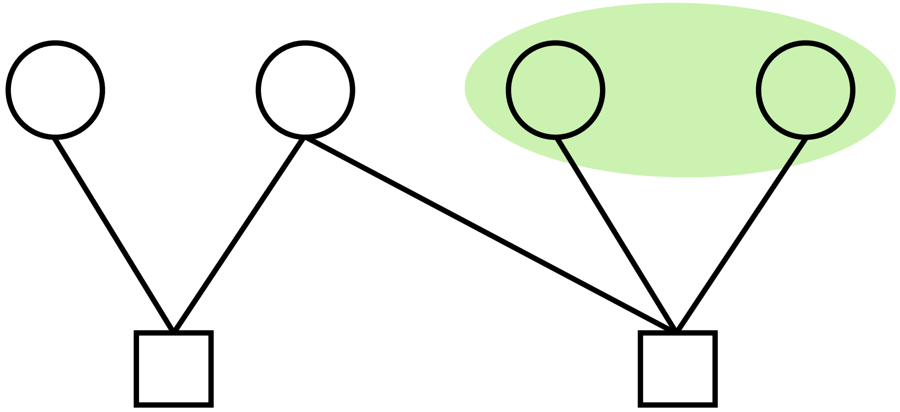
- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



Region-Based Approximation

(Definitions)

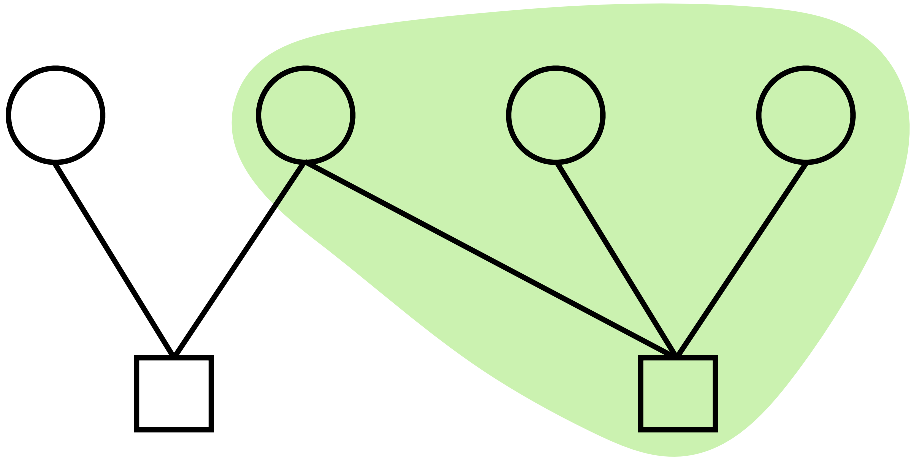
- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



Region-Based Approximation

(Definitions)

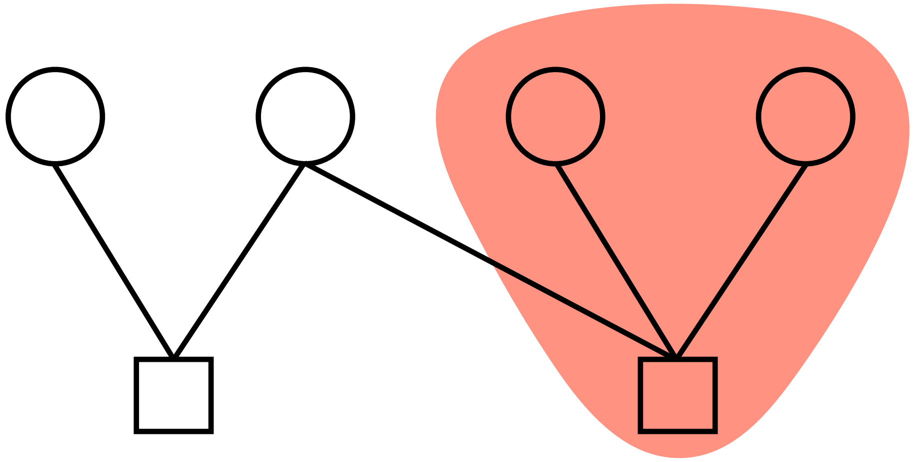
- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



Region-Based Approximation

(Definitions)

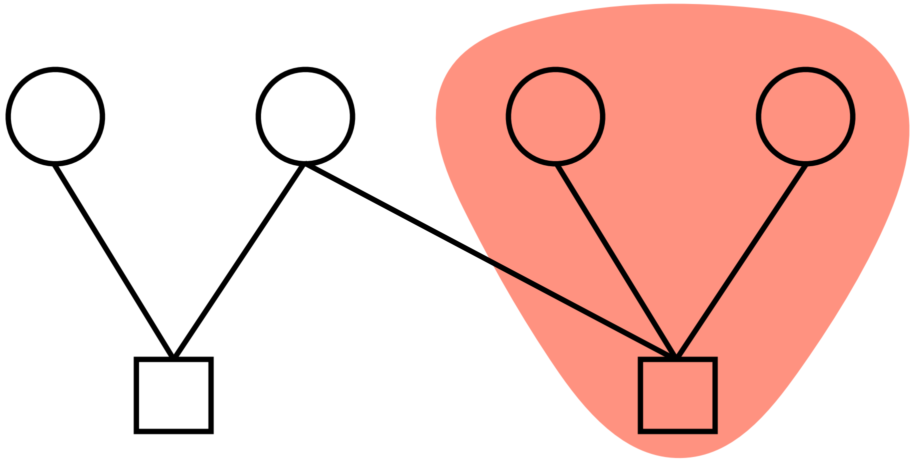
- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



Region-Based Approximation

(Definitions)

- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



- Associated quantities of a region:

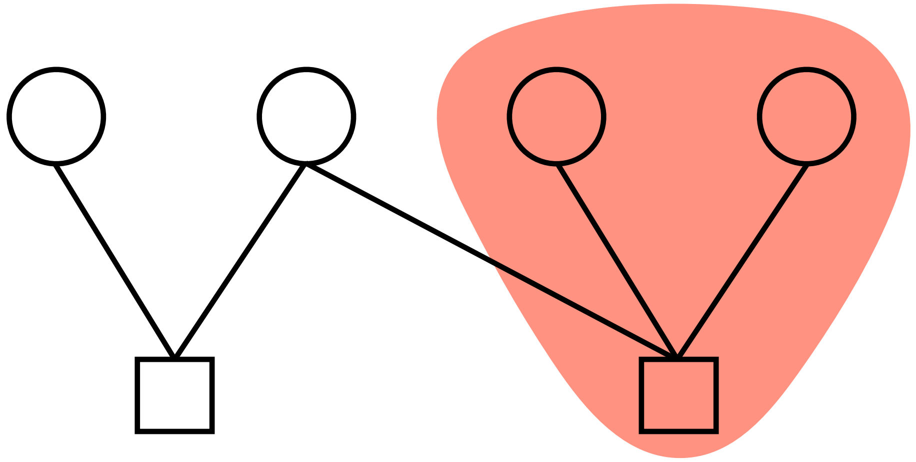
Region Energy

$$E_R(\boldsymbol{x}_R) \triangleq -\sum_{a \in A_R} \log f_a(\boldsymbol{x}_a)$$

Region-Based Approximation

(Definitions)

- A region R of a factor graph consists of V_R and A_R such that: if $a \in A_R \Rightarrow N(a) \in V_R$



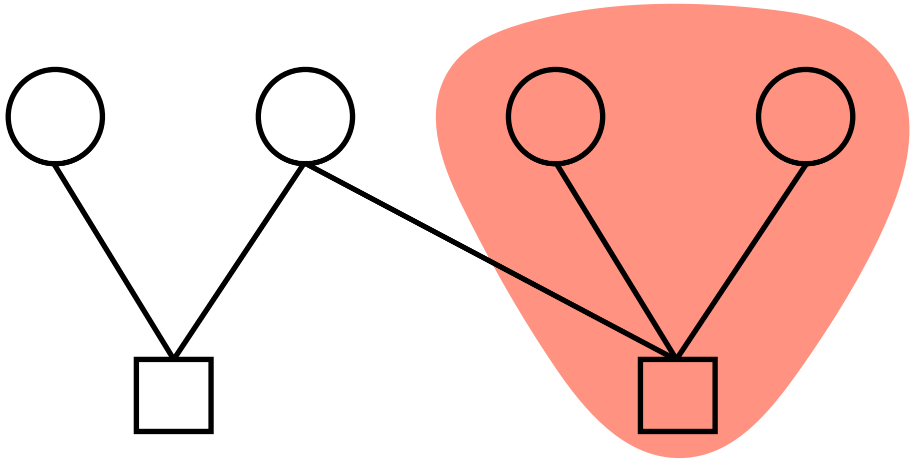
- Associated quantities of a region:

Region Energy	Region Entropy	Region Average Energy
$E_R(\boldsymbol{x}_R) \triangleq -\sum_{a \in A_R} \log f_a(\boldsymbol{x}_a)$	$H_R(b_R) \triangleq -\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) \log b_R(\boldsymbol{x}_R)$	$U_R(b_R) \triangleq \sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) E_R(\boldsymbol{x}_R)$

Region-Based Approximation

(Definitions)

- A region R of a factor graph consists of V_R and A_R such that:
  if $a \in A_R \Rightarrow N(a) \in V_R$



- Associated quantities of a region:

Region Energy

$$E_R(\boldsymbol{x}_R) \triangleq -\sum_{a \in A_R} \log f_a(\boldsymbol{x}_a)$$

Region Entropy

$$H_R(b_R) \triangleq -\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) \log b_R(\boldsymbol{x}_R)$$

Region Average Energy

$$U_R(b_R) \triangleq \sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) E_R(\boldsymbol{x}_R)$$

Region (Gibbs) Free Energy

$$F_R(b_R) \triangleq U_R(b_R) - H_R(b_R)$$

Region-Based Approximation

- Region-based (approximate) entropy:

$$H_{\mathcal{R}}(\{b_R\}) \triangleq \sum_{R\in\mathcal{R}} c_R H_R(b_R)$$

- Region-based average energy:

$$U_{\mathcal{R}}(\{b_R\}) \triangleq \sum_{R\in\mathcal{R}} c_R U_R(b_R)$$

- Region-based (Gibbs) free energy:

$$F_{\mathcal{R}}(\{b_R\}) \triangleq U_{\mathcal{R}}(\{b_R\}) - H_{\mathcal{R}}(\{b_R\})$$

Region-Based Approximation

- Region-based (approximate) entropy:

$$H_{\mathcal{R}}(\{b_R\}) \triangleq \sum_{R \in \mathcal{R}} c_R H_R(b_R)$$

- Region-based average energy:

$$U_{\mathcal{R}}(\{b_R\}) \triangleq \sum_{R \in \mathcal{R}} c_R U_R(b_R)$$

Counting numbers

- Region-based (Gibbs) free energy:

$$F_{\mathcal{R}}(\{b_R\}) \triangleq U_{\mathcal{R}}(\{b_R\}) - H_{\mathcal{R}}(\{b_R\})$$

Valid Region-Based Approximation

- Definition: A set of regions $\mathcal{R}$ and associated counting numbers c_R give a valid approximation if:

$$\sum_{R \in \mathcal{R}} c_R I_{A_R}(a) = \sum_{R \in \mathcal{R}} c_R I_{V_R}(i) = 1, \qquad \forall i \in V \text{ and } \forall a \in A$$

- Why valid region-based approximation?
    - If $b_R(\boldsymbol{x}) = p_R(\boldsymbol{x}) \Rightarrow U = U_{\mathcal{R}}(\{b_R\})$
    - In general $H \neq H_{\mathcal{R}}(\{b_R\})$ but H is equal to $H_{\mathcal{R}}(\{b_R\})$ up to total number of degrees of freedom in the system

Valid Region-Based Approximation

- Definition: A set of regions $\mathcal{R}$ and associated counting numbers c_R give a valid approximation if:

$$\sum_{R\in\mathcal{R}} c_R I_{A_R}(a) = \sum_{R\in\mathcal{R}} c_R I_{V_R}(i) = 1, \qquad \forall i \in V \text{ and } \forall a \in A$$

- Why valid region-based approximation?
  - ✓ If $b_R(\boldsymbol{x}) = p_R(\boldsymbol{x}) \Rightarrow U = U_{\mathcal{R}}(\{b_R\})$
  - In general $H \neq H_{\mathcal{R}}(\{b_R\})$ but H is equal to $H_{\mathcal{R}}(\{b_R\})$ up to total number of degrees of freedom in the system

Valid Region-Based Approximation

- Definition: A set of regions $\mathcal{R}$ and associated counting numbers c_R give a valid approximation if:

$$\sum_{R\in\mathcal{R}} c_R I_{A_R}(a) = \sum_{R\in\mathcal{R}} c_R I_{V_R}(i) = 1, \qquad \forall i \in V \text{ and } \forall a \in A$$

- Why valid region-based approximation?
  - ✓ If $b_R(\boldsymbol{x}) = p_R(\boldsymbol{x}) \Rightarrow U = U_{\mathcal{R}}(\{b_R\})$
  - ✗ In general $H \neq H_{\mathcal{R}}(\{b_R\})$ but H is equal to $H_{\mathcal{R}}(\{b_R\})$ up to total number of degrees of freedom in the system

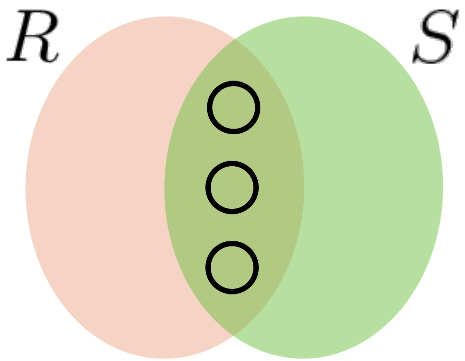
Region-Based Approximation
(Constraints on Beliefs)

1. Normalization: $\forall R \in \mathcal{R},\ b_R(\boldsymbol{x}_R)$ forms a probability function:

$$\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) = 1$$

2. Local consistency: if the set of variable nodes $W \subseteq R \cap S$:

$$\sum_{\boldsymbol{x}_R \setminus \boldsymbol{x}_W} b_R(\boldsymbol{x}_R) = \sum_{\boldsymbol{x}_S \setminus \boldsymbol{x}_W} b_S(\boldsymbol{x}_S)$$



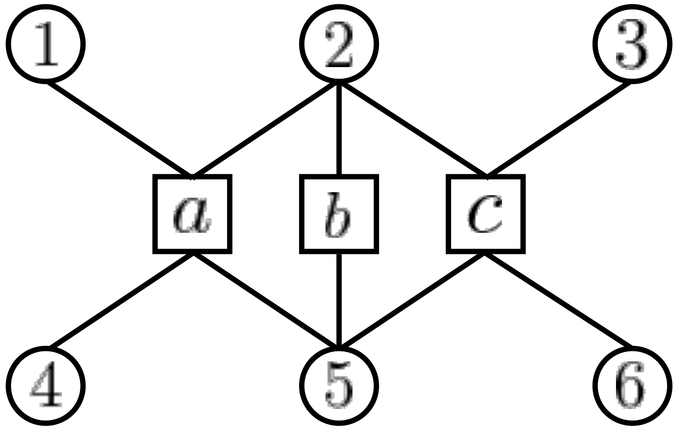
3. Inequality: $0 \leq b_R(\boldsymbol{x}_R) \leq 1$

The above expressions give a set of local constraints!

A Special Case:
Bethe Approximation
and
Recovering BP

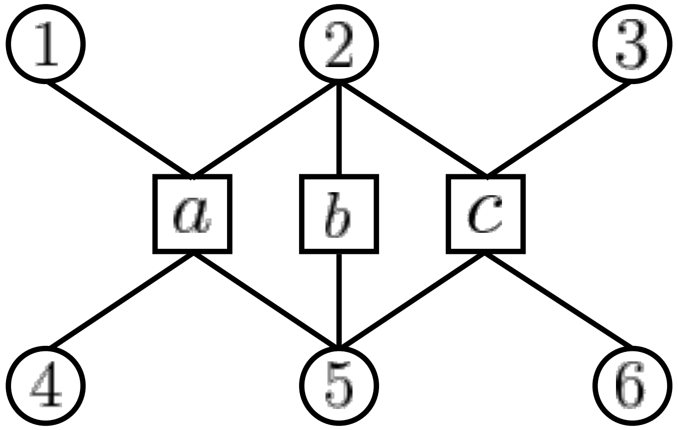
Bethe Approximation

- Two types of regions, large and small: $\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$
- n regions in $\mathcal{R}_S$ each contains one variable node
- m regions in $\mathcal{R}_L$ each contains one factor node and the neighboring variable nodes



Bethe Approximation

- Two types of regions, large and small: $\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$
- n regions in $\mathcal{R}_S$ each contains one variable node
- m regions in $\mathcal{R}_L$ each contains one factor node and the neighboring variable nodes

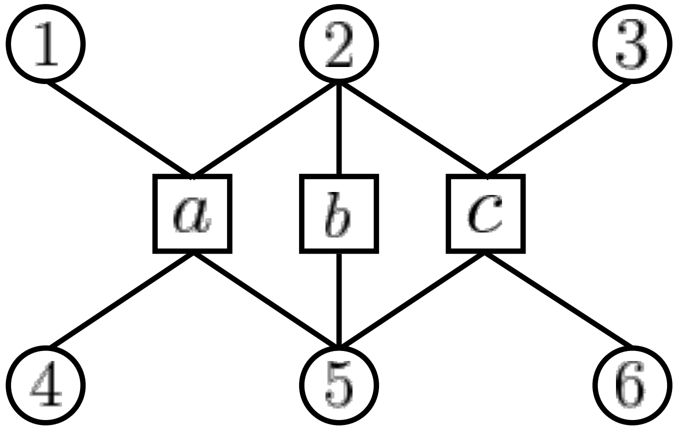


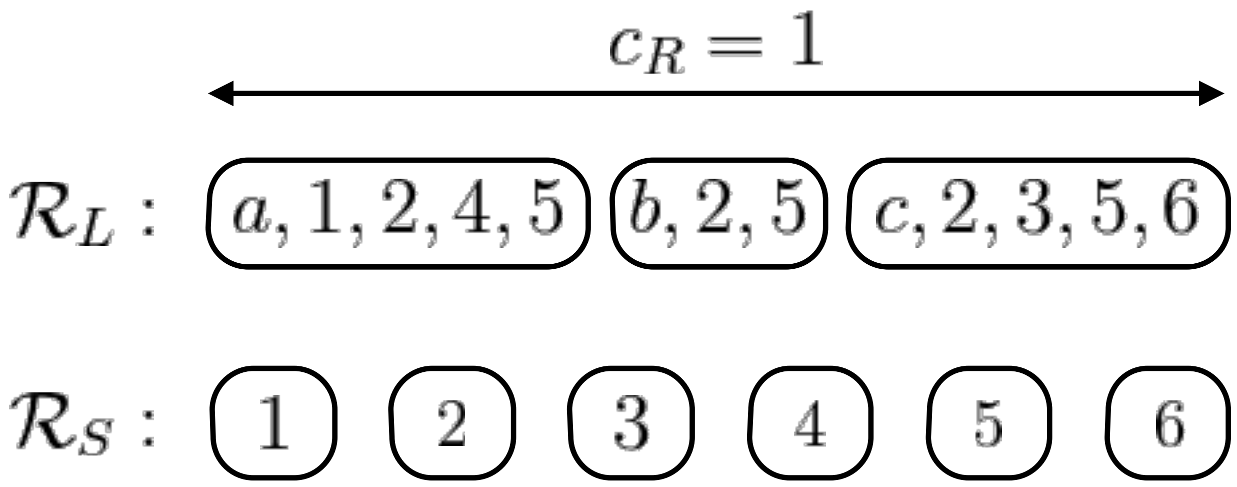
$\mathcal{R}_L$: $\boxed{a, 1, 2, 4, 5}$ $\boxed{b, 2, 5}$ $\boxed{c, 2, 3, 5, 6}$

$\mathcal{R}_S$: $\boxed{1}$ $\boxed{2}$ $\boxed{3}$ $\boxed{4}$ $\boxed{5}$ $\boxed{6}$

Bethe Approximation

- Two types of regions, large and small: $\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$
- n regions in $\mathcal{R}_S$ each contains one variable node
- m regions in $\mathcal{R}_L$ each contains one factor node and the neighboring variable nodes





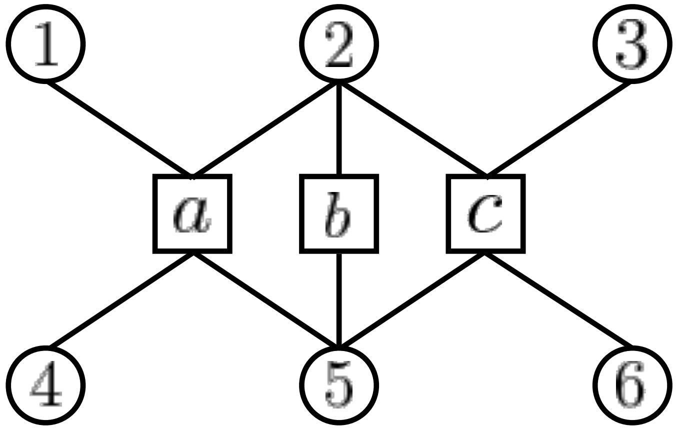
$c_R = 1$

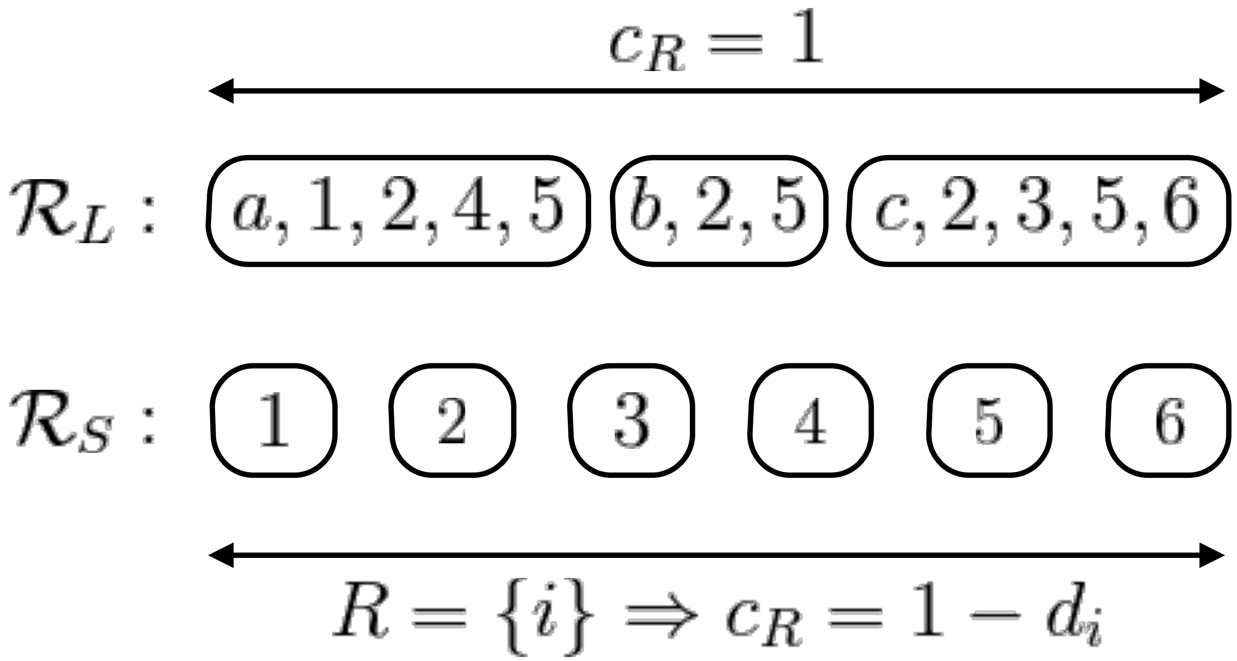
$\mathcal{R}_L :$ $a, 1, 2, 4, 5$ $b, 2, 5$ $c, 2, 3, 5, 6$

$\mathcal{R}_S :$ 1 2 3 4 5 6

Bethe Approximation

- Two types of regions, large and small: $\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$
- n regions in $\mathcal{R}_S$ each contains one variable node
- m regions in $\mathcal{R}_L$ each contains one factor node and the neighboring variable nodes





$c_R = 1$

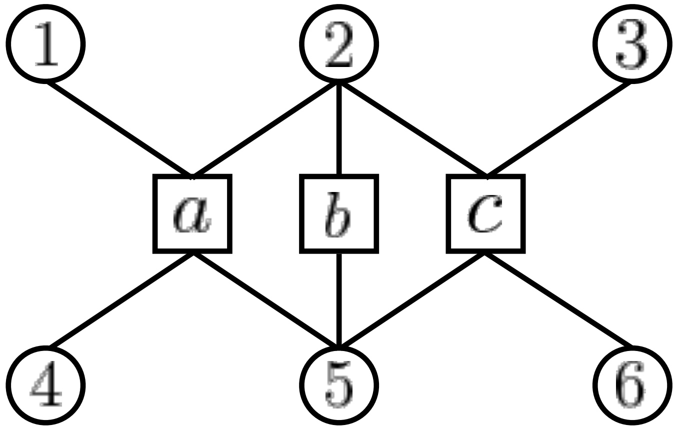
$\mathcal{R}_L :$ $a, 1, 2, 4, 5$ $b, 2, 5$ $c, 2, 3, 5, 6$

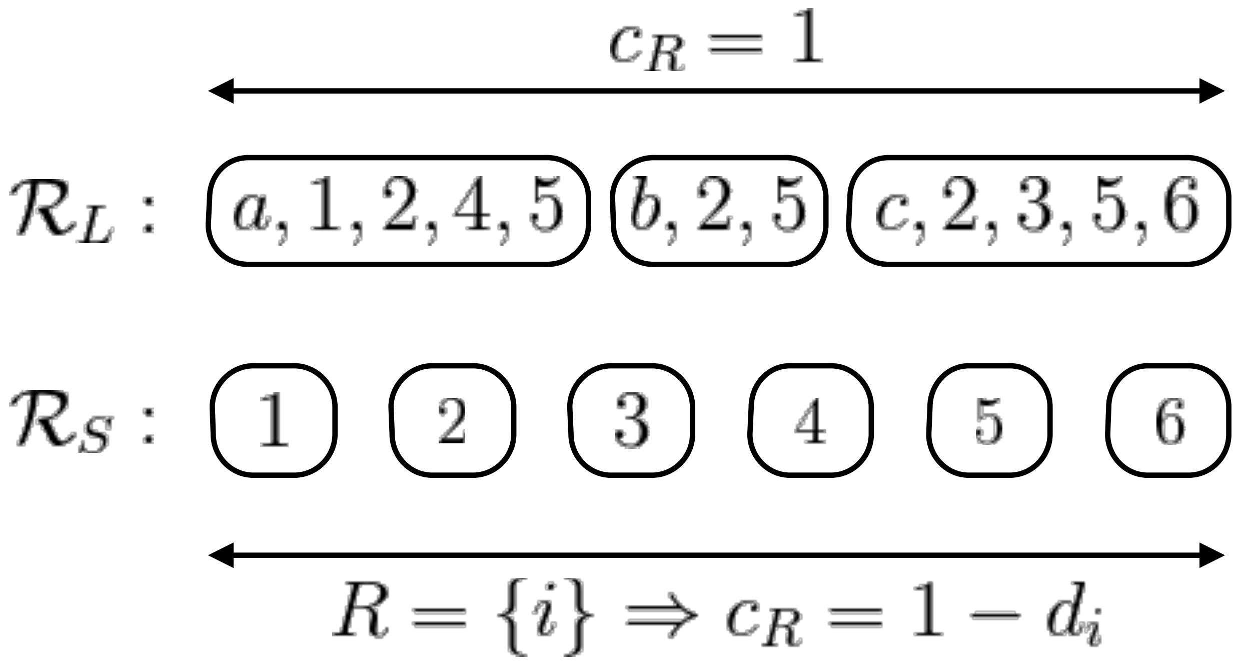
$\mathcal{R}_S :$ 1 2 3 4 5 6

$R = \{i\} \Rightarrow c_R = 1 - d_i$

Bethe Approximation

- Two types of regions, large and small: $\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$
- n regions in $\mathcal{R}_S$ each contains one variable node
- m regions in $\mathcal{R}_L$ each contains one factor node and the neighboring variable nodes





$c_R = 1$

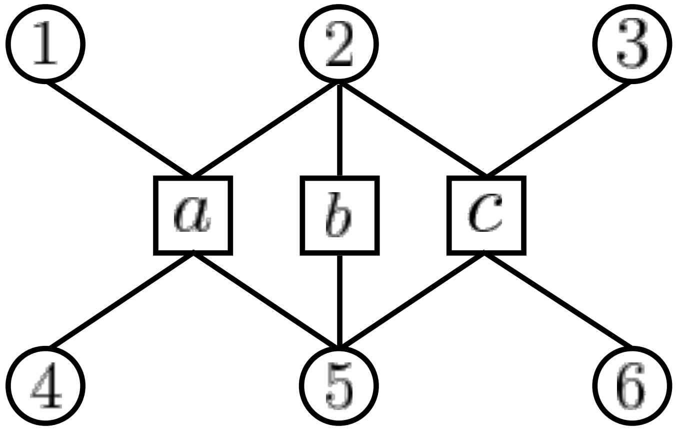
$\mathcal{R}_L :$ $a, 1, 2, 4, 5$ $b, 2, 5$ $c, 2, 3, 5, 6$

$\mathcal{R}_S :$ 1 2 3 4 5 6

$R = \{i\} \Rightarrow c_R = 1 - d_i$

Good news: this choice of counting numbers give a valid approximation for variational free energy!

Bethe Approximation



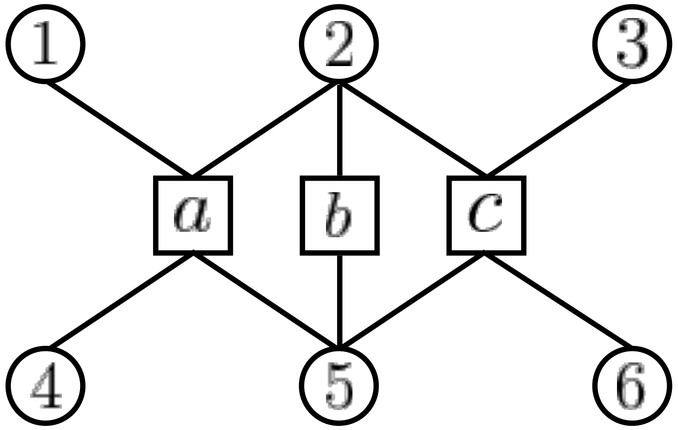
Bethe Average Energy

$$U_{\mathrm{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log f_a(\boldsymbol{x}_a)$$

Bethe Entropy

$$H_{\mathrm{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log b_a(\boldsymbol{x}_a) + \sum_{i \in V} (d_i - 1) \sum_{x_i} b_i(x_i) \log b_i(x_i)$$

Bethe Approximation



Bethe Average Energy

$$U_{\mathrm{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log f_a(\boldsymbol{x}_a)$$

Bethe Entropy

$$H_{\mathrm{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log b_a(\boldsymbol{x}_a) + \sum_{i \in V} (d_i - 1) \sum_{x_i} b_i(x_i) \log b_i(x_i)$$

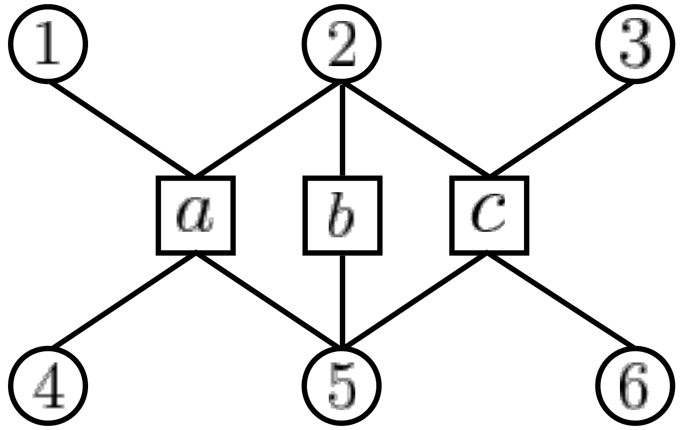
If the factor graph has no cycle → Bethe approximation is exact:

$H_{\mathrm{Bethe}} = H \;$ if $\; b(\boldsymbol{x}) = p(\boldsymbol{x})$

Bethe Approximation

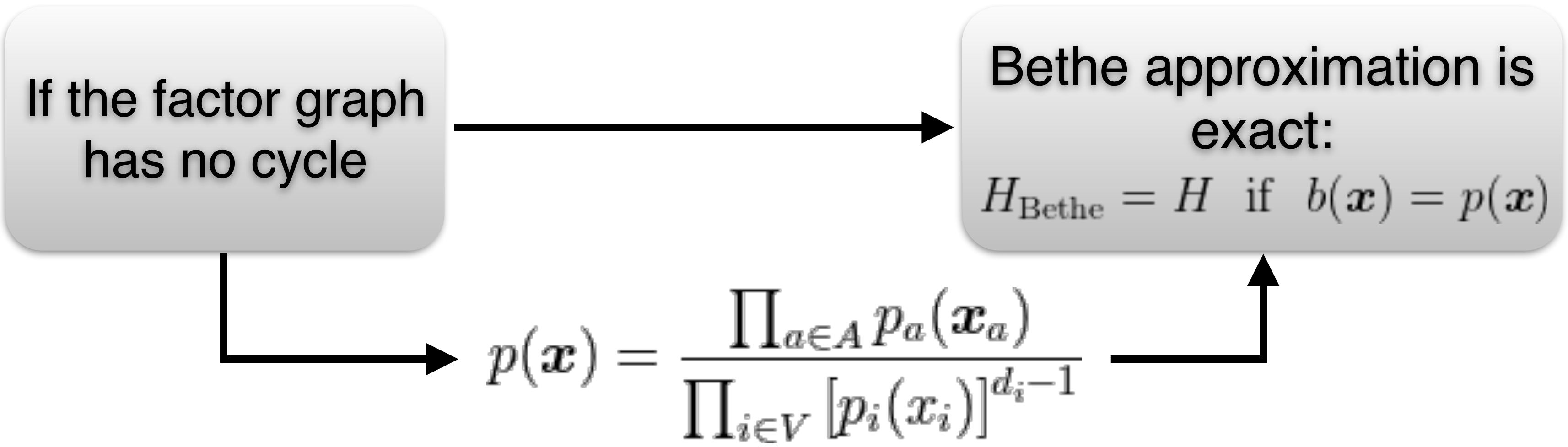
Bethe Average Energy

$$U_{\text{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log f_a(\boldsymbol{x}_a)$$



Bethe Entropy

$$H_{\text{Bethe}} = -\sum_{a \in A} \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) \log b_a(\boldsymbol{x}_a) + \sum_{i \in V} (d_i - 1) \sum_{x_i} b_i(x_i) \log b_i(x_i)$$



If the factor graph has no cycle

Bethe approximation is exact:

$$H_{\text{Bethe}} = H \quad \text{if} \quad b(\boldsymbol{x}) = p(\boldsymbol{x})$$

$$p(\boldsymbol{x}) = \frac{\prod_{a \in A} p_a(\boldsymbol{x}_a)}{\prod_{i \in V} [p_i(x_i)]^{d_i - 1}}$$

Bethe Approximation

(Constraints on Beliefs)

- Constraints:
  - Normalization: $\sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = \sum_{x_i} b_i(x_i) = 1, \quad \forall i \in V \text{ and } \forall a \in A$
  - Consistency: $\sum_{\boldsymbol{x}_a \setminus x_i} b_a(\boldsymbol{x}_a) = b_i(x_i), \quad \forall a \in A \text{ and } \forall i \in N(a)$
  - Inequality: $0 \leq b(\boldsymbol{x}_a) \leq 1, \quad 0 \leq b_i(x_i) \leq 1, \quad \forall a \in A \text{ and } \forall i \in V$

Bethe Approximation

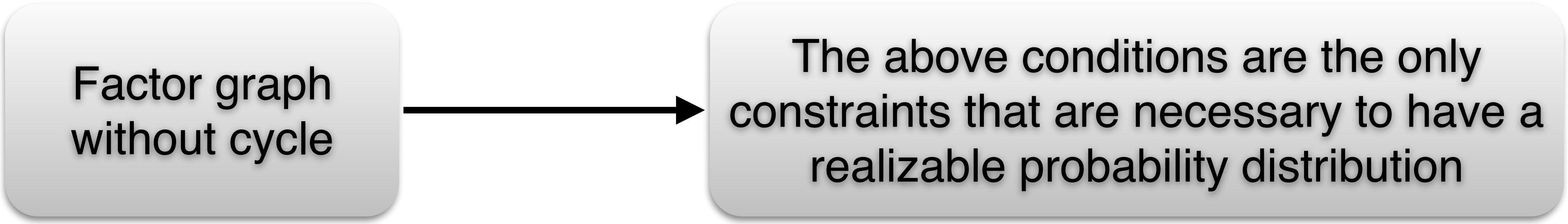
(Constraints on Beliefs)

- Constraints:
  - Normalization: $\sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = \sum_{x_i} b_i(x_i) = 1, \quad \forall i \in V \text{ and } \forall a \in A$
  - Consistency: $\sum_{\boldsymbol{x}_a \backslash x_i} b_a(\boldsymbol{x}_a) = b_i(x_i), \quad \forall a \in A \text{ and } \forall i \in N(a)$
  - Inequality: $0 \leq b(\boldsymbol{x}_a) \leq 1, \quad 0 \leq b_i(x_i) \leq 1, \quad \forall a \in A \text{ and } \forall i \in V$
- Bad news:
  - The above constraints do not necessarily lead to a probability distribution over $\boldsymbol{x}$!
  - We me have negative entropy!

Bethe Approximation

(Constraints on Beliefs)

- Constraints:
  - Normalization: $\sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = \sum_{x_i} b_i(x_i) = 1, \quad \forall i \in V \text{ and } \forall a \in A$
  - Consistency: $\sum_{\boldsymbol{x}_a \setminus x_i} b_a(\boldsymbol{x}_a) = b_i(x_i), \quad \forall a \in A \text{ and } \forall i \in N(a)$
  - Inequality: $0 \leq b(\boldsymbol{x}_a) \leq 1, \quad 0 \leq b_i(x_i) \leq 1, \quad \forall a \in A \text{ and } \forall i \in V$
- Bad news:
  - The above constraints do not necessarily lead to a probability distribution over $\boldsymbol{x}$!
  - We me have negative entropy!

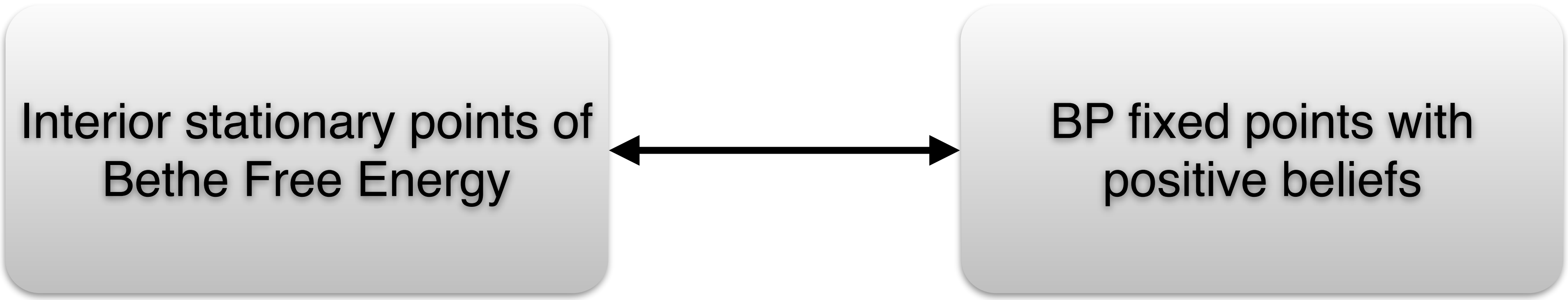


Factor graph without cycle

The above conditions are the only constraints that are necessary to have a realizable probability distribution

Connection Between Bethe Approximation and BP

- Theorem:

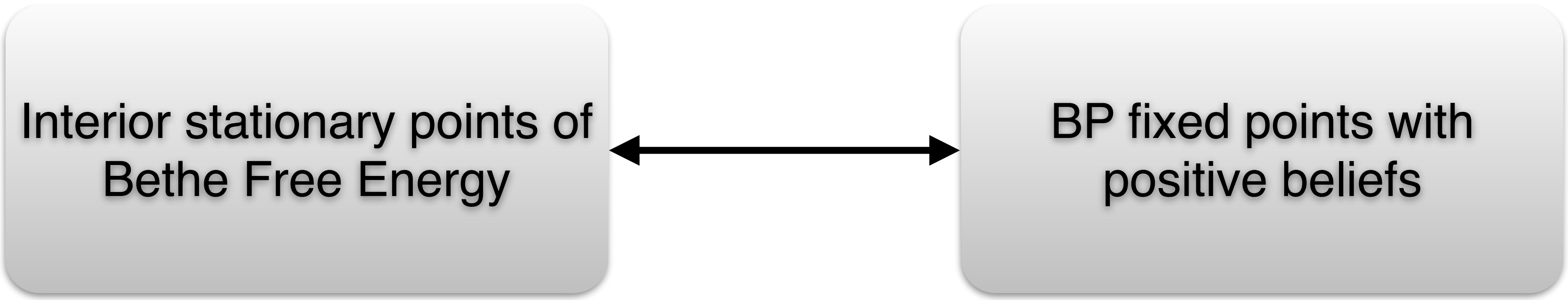


Interior stationary points of Bethe Free Energy

BP fixed points with positive beliefs

Connection Between Bethe Approximation and BP

- Theorem:



Interior stationary points of Bethe Free Energy

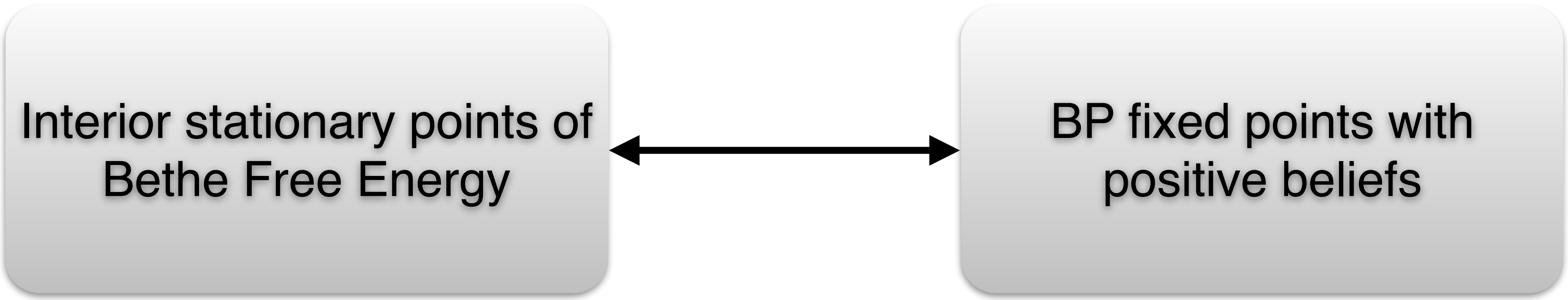
BP fixed points with positive beliefs

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

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

Connection Between Bethe Approximation and BP

- Theorem:



Interior stationary points of Bethe Free Energy

BP fixed points with positive beliefs

$$\min_b F_{\text{Bethe}} = \min_b \left[U_{\text{Bethe}} - H_{\text{Bethe}}\right]$$

$$\text{s.t.} \quad \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = 1$$

$$\sum_{\boldsymbol{x}_a \setminus x_i} b_a(\boldsymbol{x}_a) = b_i(x_i)$$

$$\sum_{x_i} b_i(x_i) = 1$$

$$0 \leq b_a(\boldsymbol{x}_a) \leq 1$$

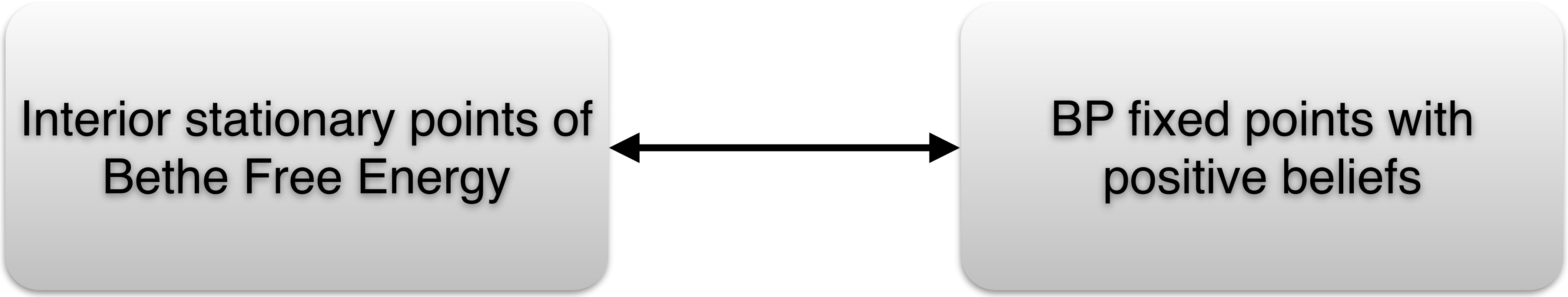
$$0 \leq b_i(x_i) \leq 1$$

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

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

Connection Between Bethe Approximation and BP

- Theorem:



Interior stationary points of Bethe Free Energy

BP fixed points with positive beliefs

$$\min_b F_{\text{Bethe}} = \min_b \left[U_{\text{Bethe}} - H_{\text{Bethe}}\right]$$

$$\text{s.t.} \quad \left.\begin{aligned} &\sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = 1 \\ &\sum_{\boldsymbol{x}_a \setminus x_i} b_a(\boldsymbol{x}_a) = b_i(x_i) \\ &\sum_{x_i} b_i(x_i) = 1 \end{aligned}\right\} \longleftarrow \boxed{\text{Leads to the interior stationary points}}$$

$$0 \leq b_a(\boldsymbol{x}_a) \leq 1$$

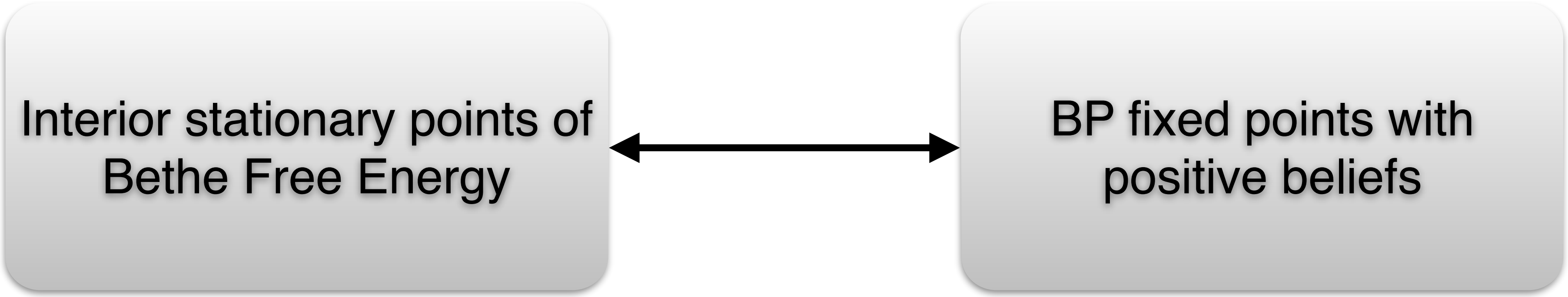
$$0 \leq b_i(x_i) \leq 1$$

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

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

Connection Between Bethe Approximation and BP

- Theorem:



Interior stationary points of Bethe Free Energy

BP fixed points with positive beliefs

$$\min_b F_{\text{Bethe}} = \min_b [U_{\text{Bethe}} - H_{\text{Bethe}}]$$

$$\text{s.t.} \quad \sum_{\boldsymbol{x}_a} b_a(\boldsymbol{x}_a) = 1$$

$$\sum_{\boldsymbol{x}_a \setminus x_i} b_a(\boldsymbol{x}_a) = b_i(x_i)$$

$$\sum_{x_i} b_i(x_i) = 1$$

$$0 \leq b_a(\boldsymbol{x}_a) \leq 1$$

$$0 \leq b_i(x_i) \leq 1$$

Leads to the interior stationary points

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

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

Proof Idea (using Lagrange method)

- Write the Lagrangian of the Bethe optimization problem
- Take derivative of $\mathcal{L}$ and find the stationary points of F_{Bethe}
- By appropriate change of variables, connect them to BP update rule

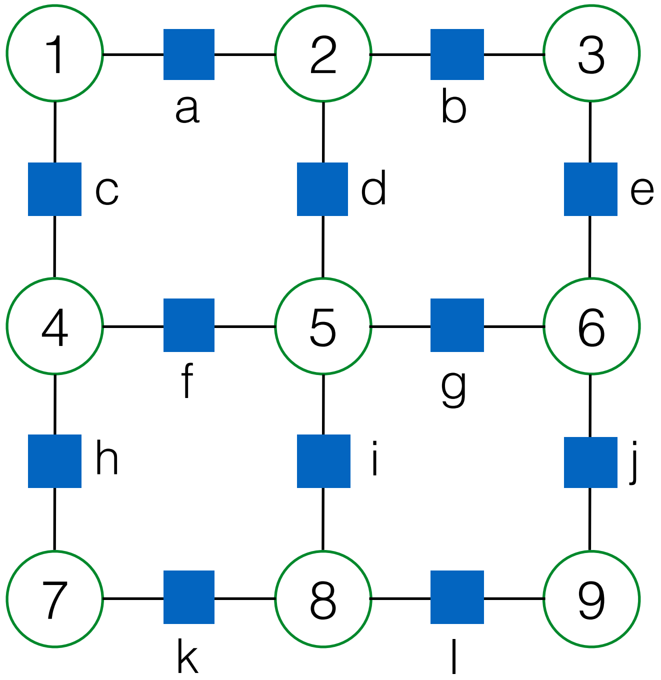
Region Graph Method
and
Generalized Belief Propagation

The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

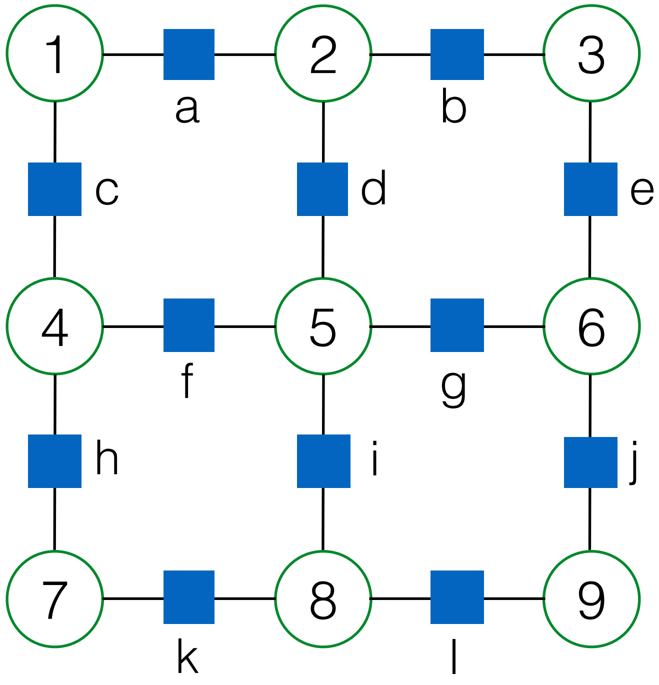
The Region Graph Method

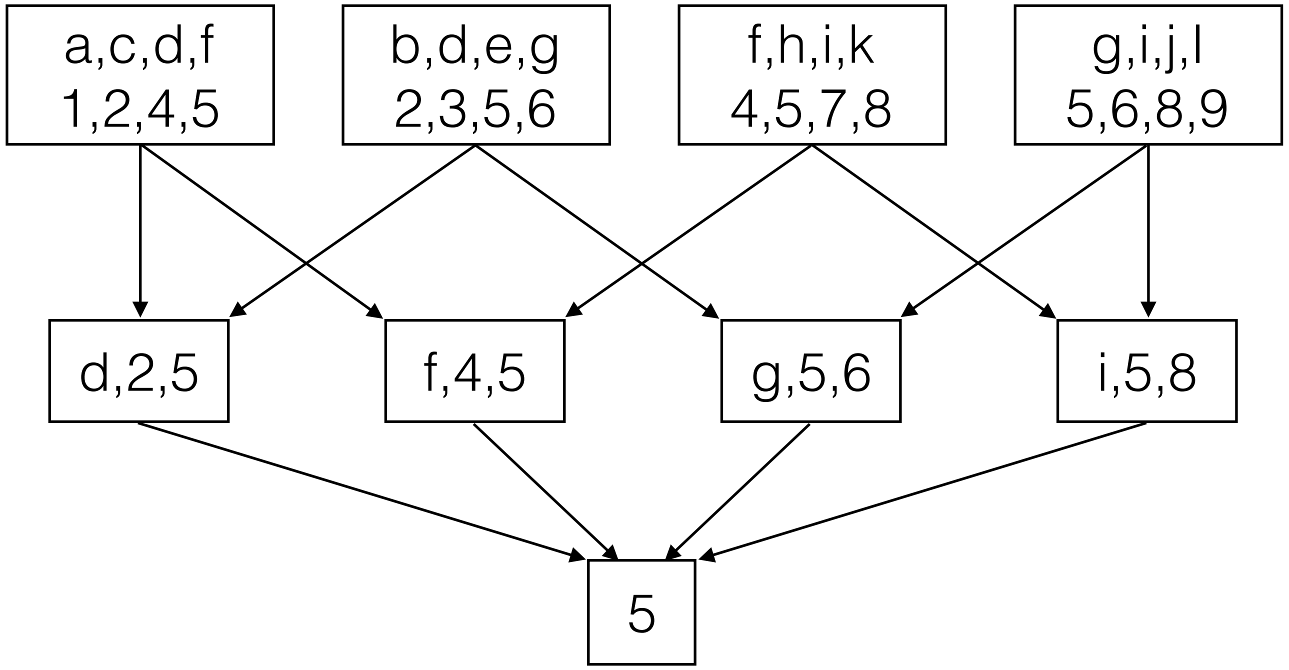
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$



The Region Graph Method

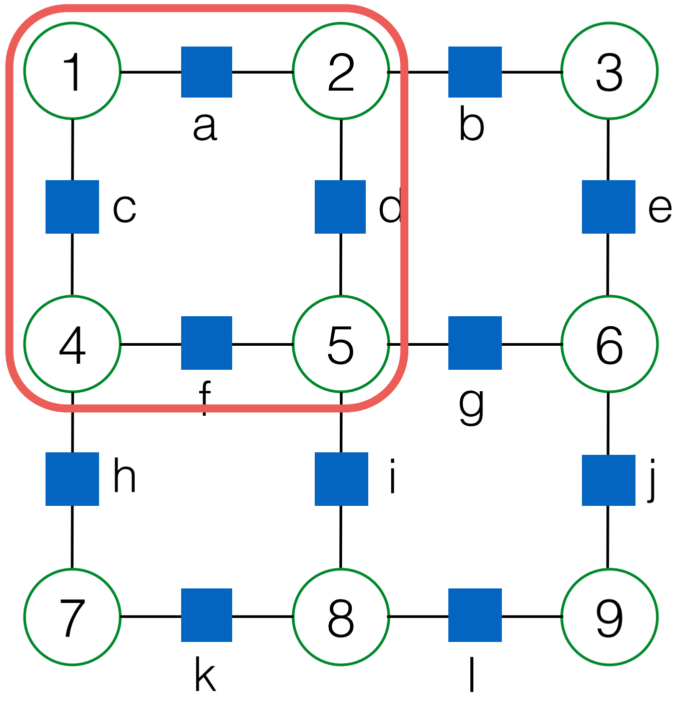
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

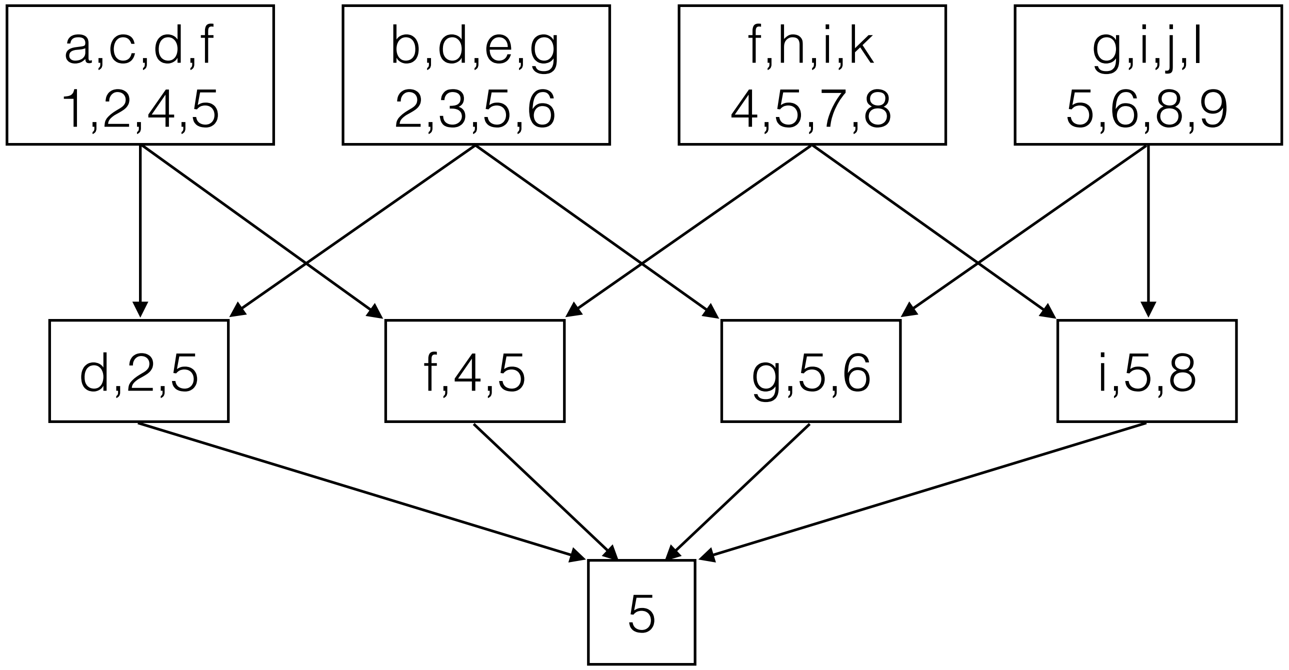




The Region Graph Method

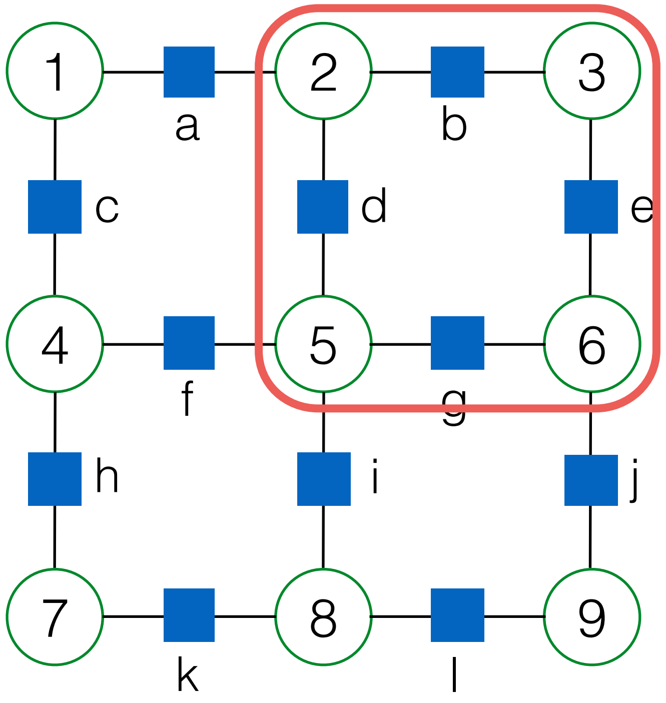
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

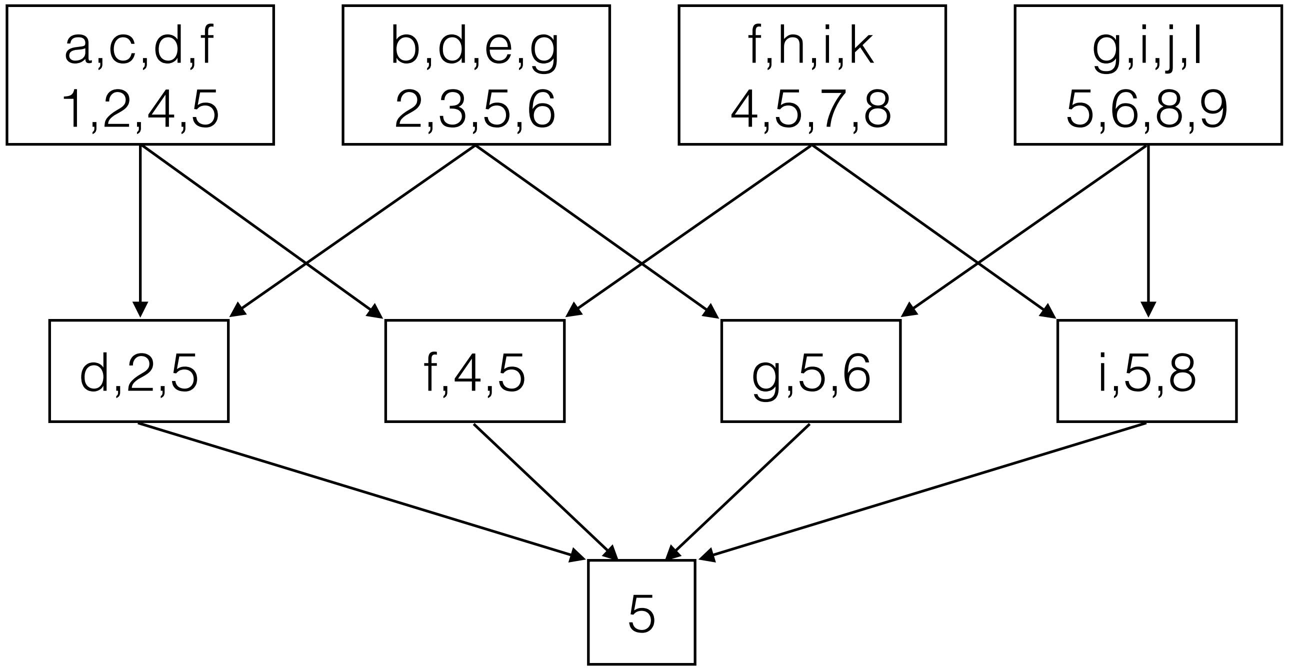




The Region Graph Method

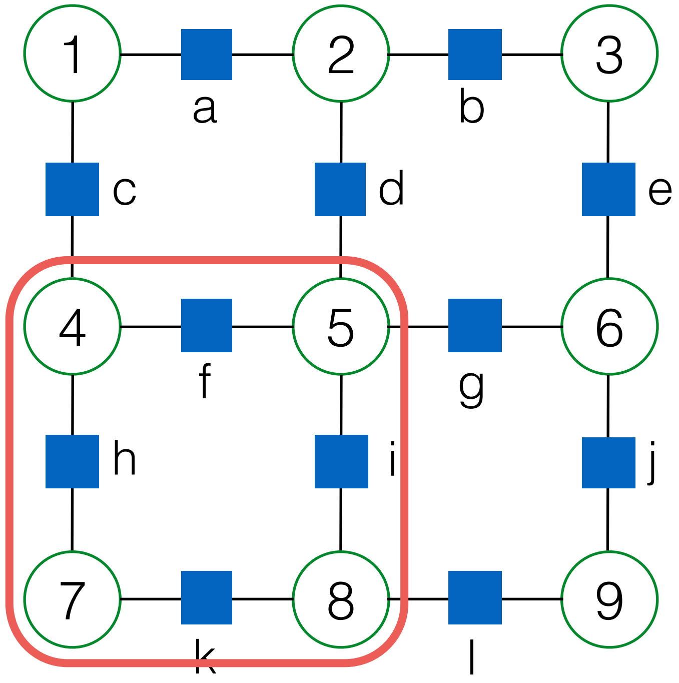
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

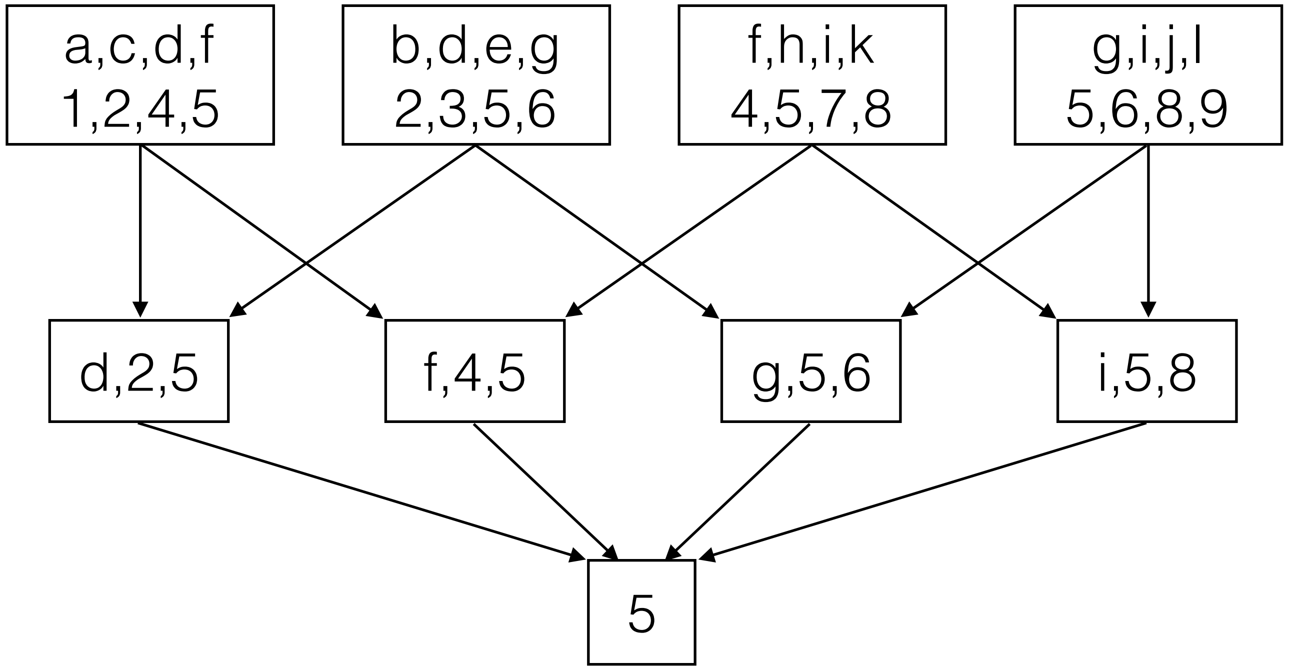




The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

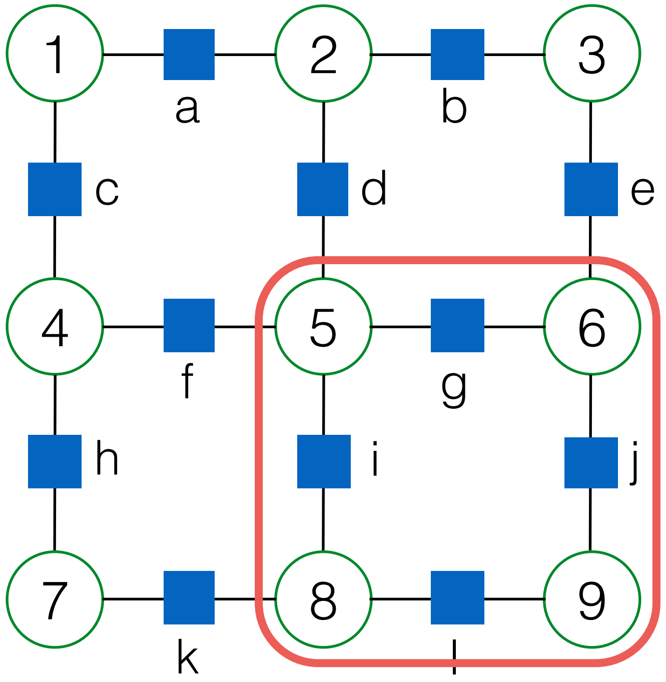
g,5,6

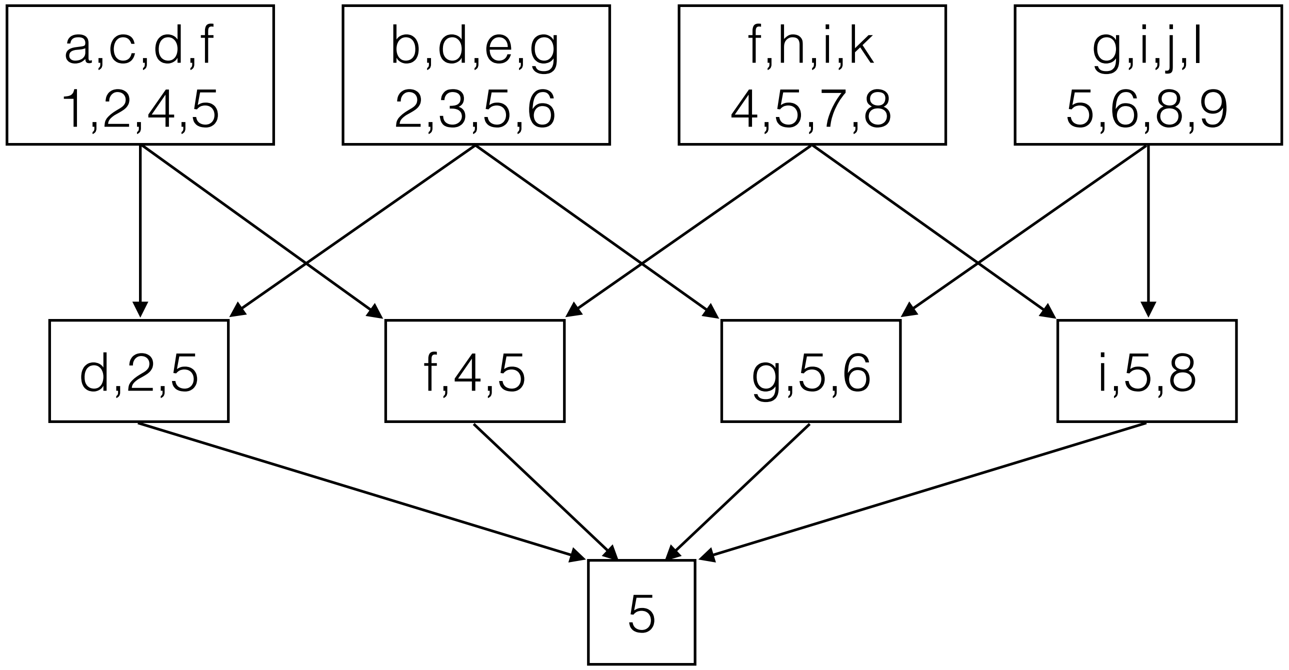
i,5,8

5

The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

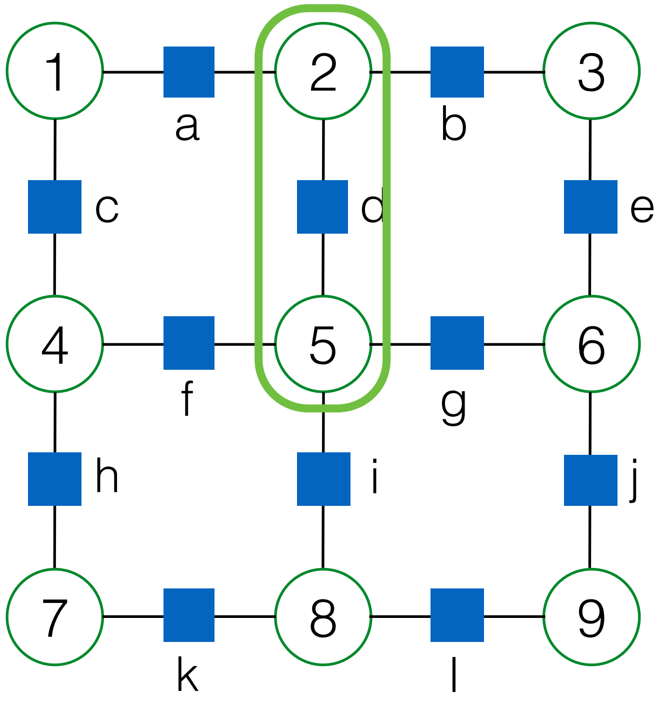
g,5,6

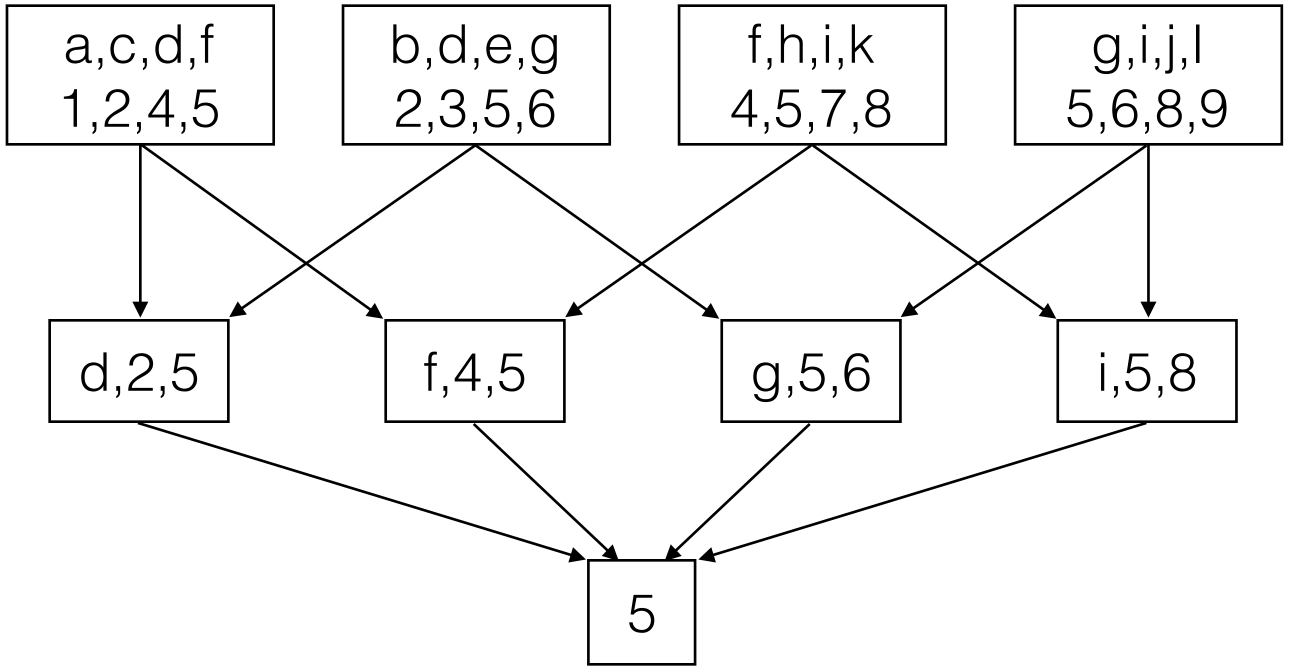
i,5,8

5

The Region Graph Method

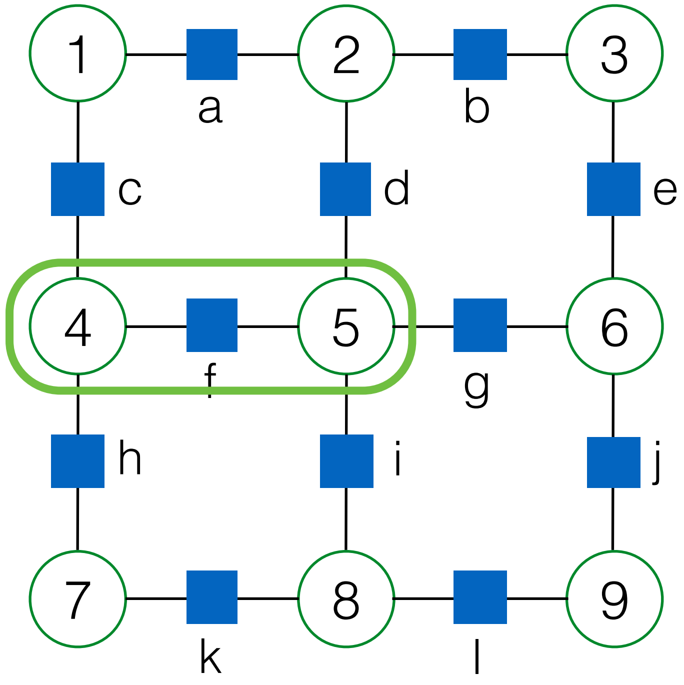
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

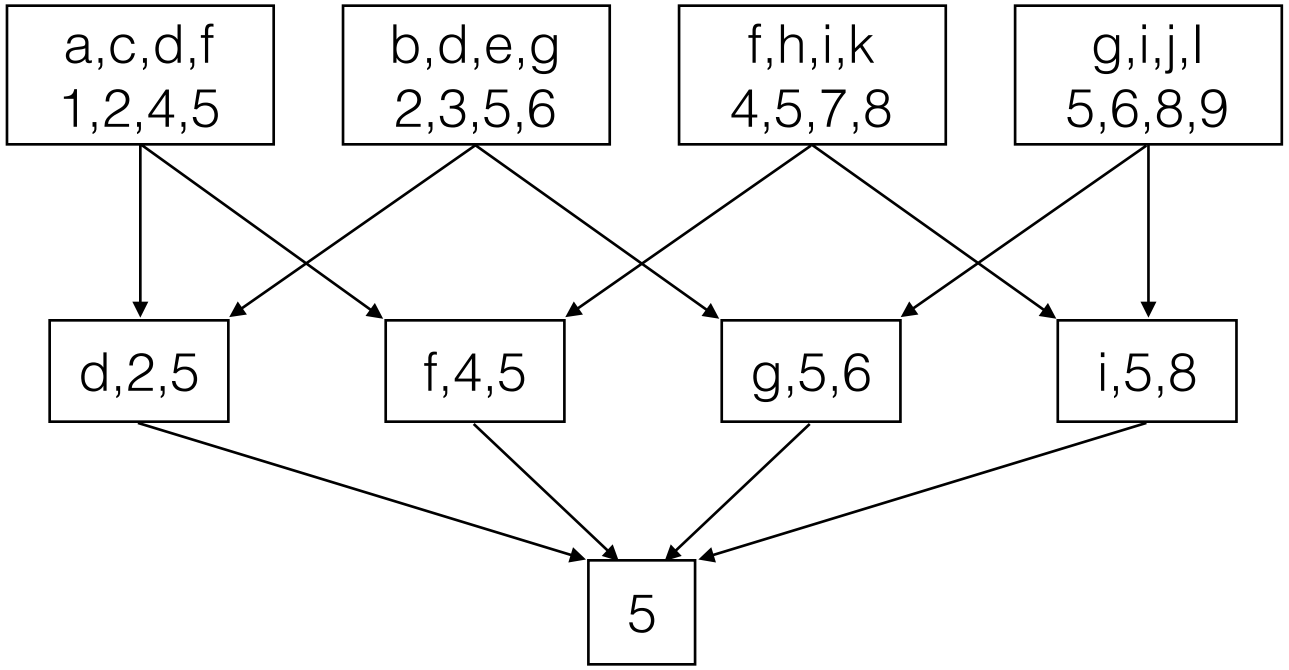




The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

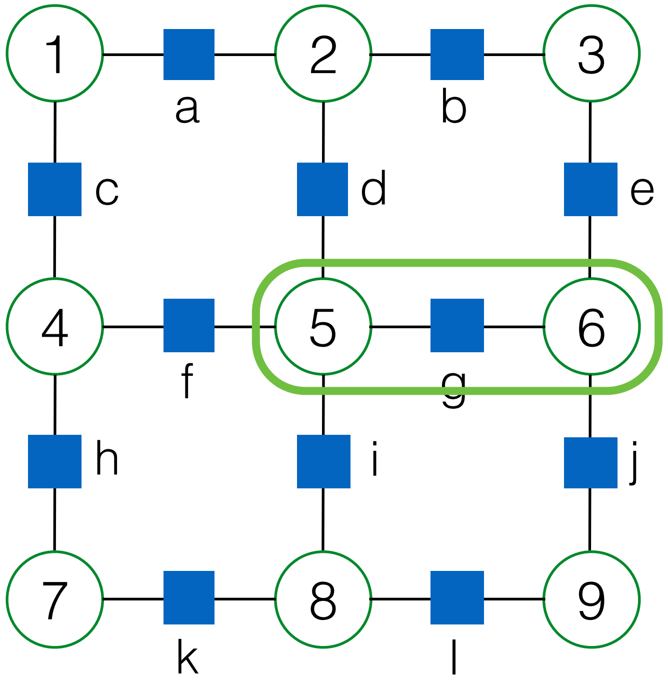
g,5,6

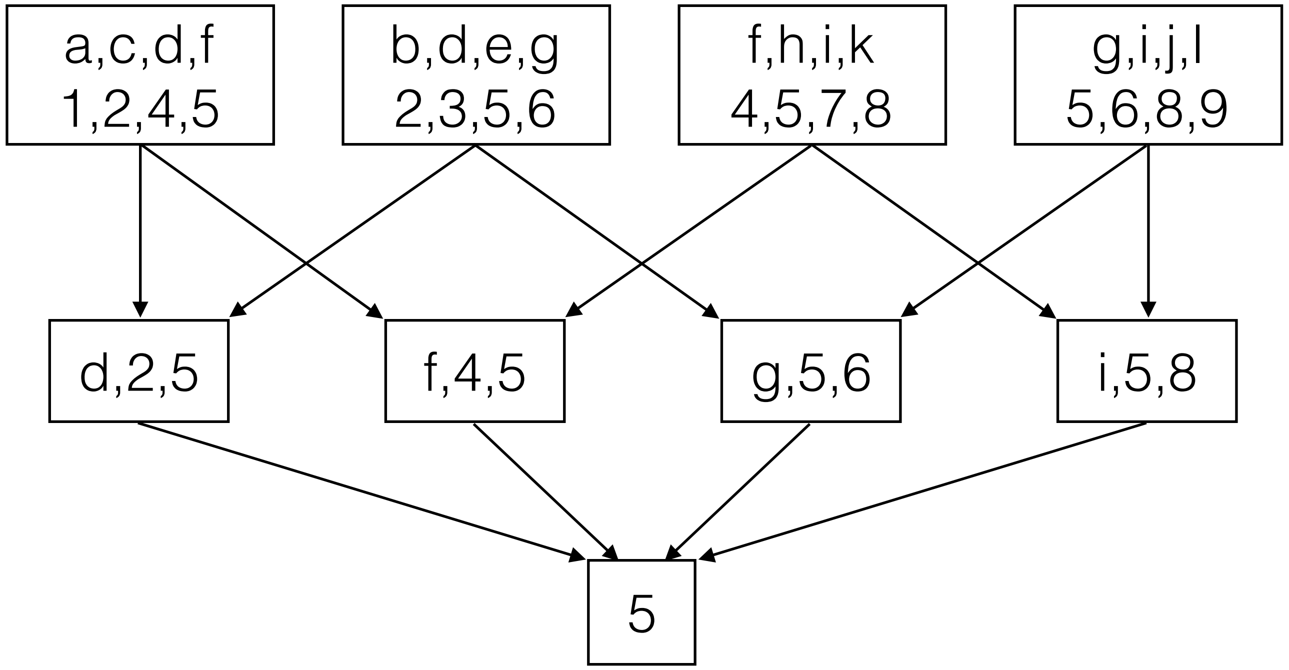
i,5,8

5

The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

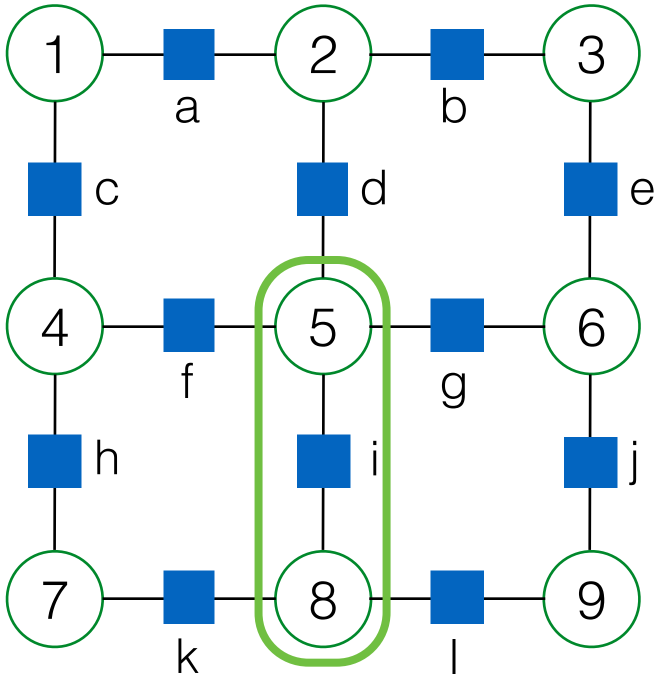
g,5,6

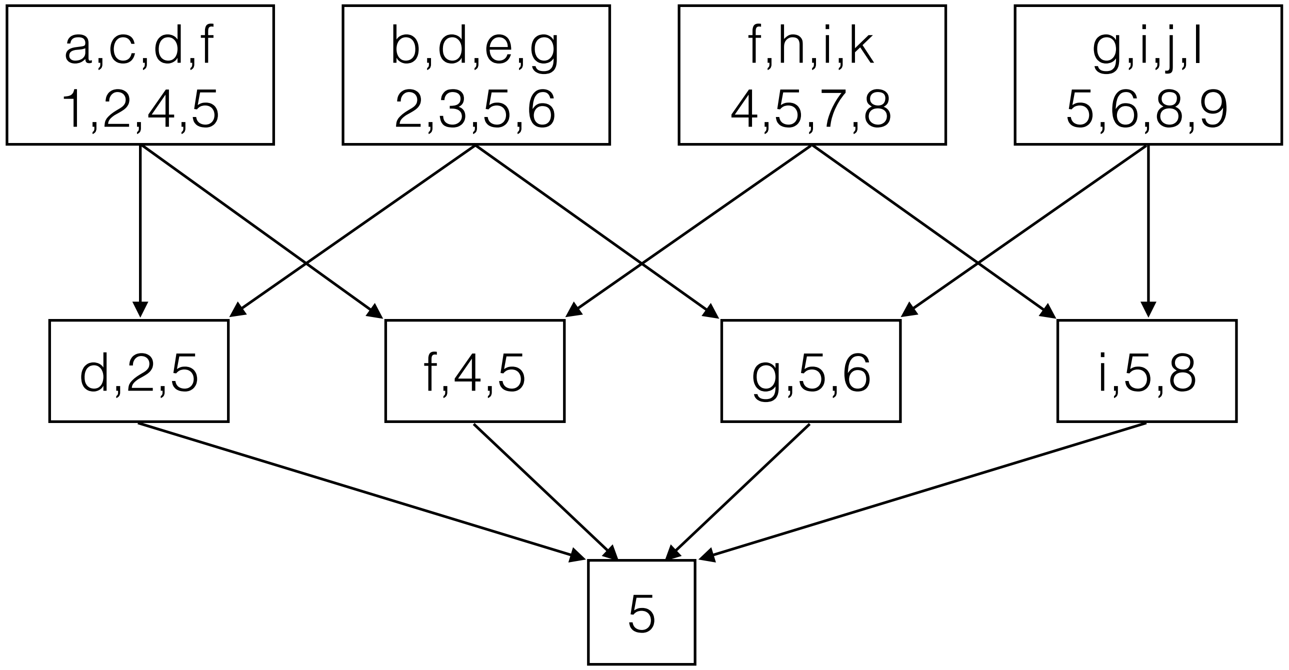
i,5,8

5

The Region Graph Method

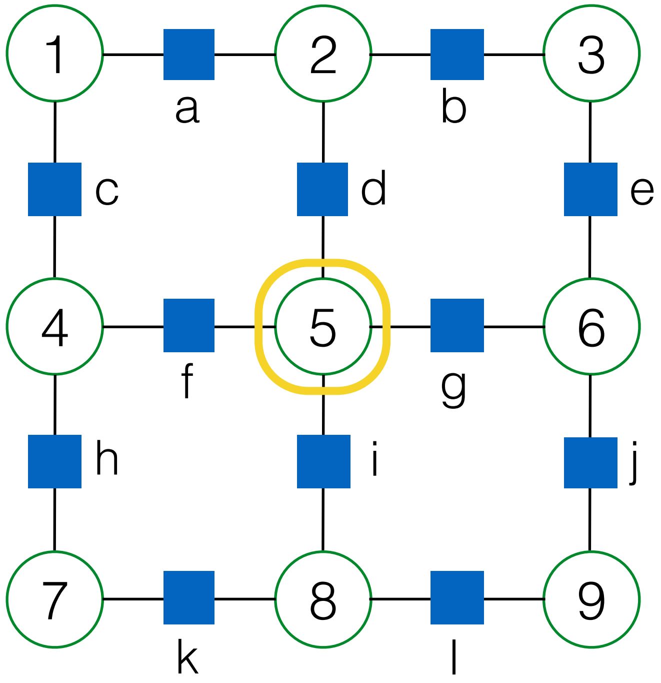
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

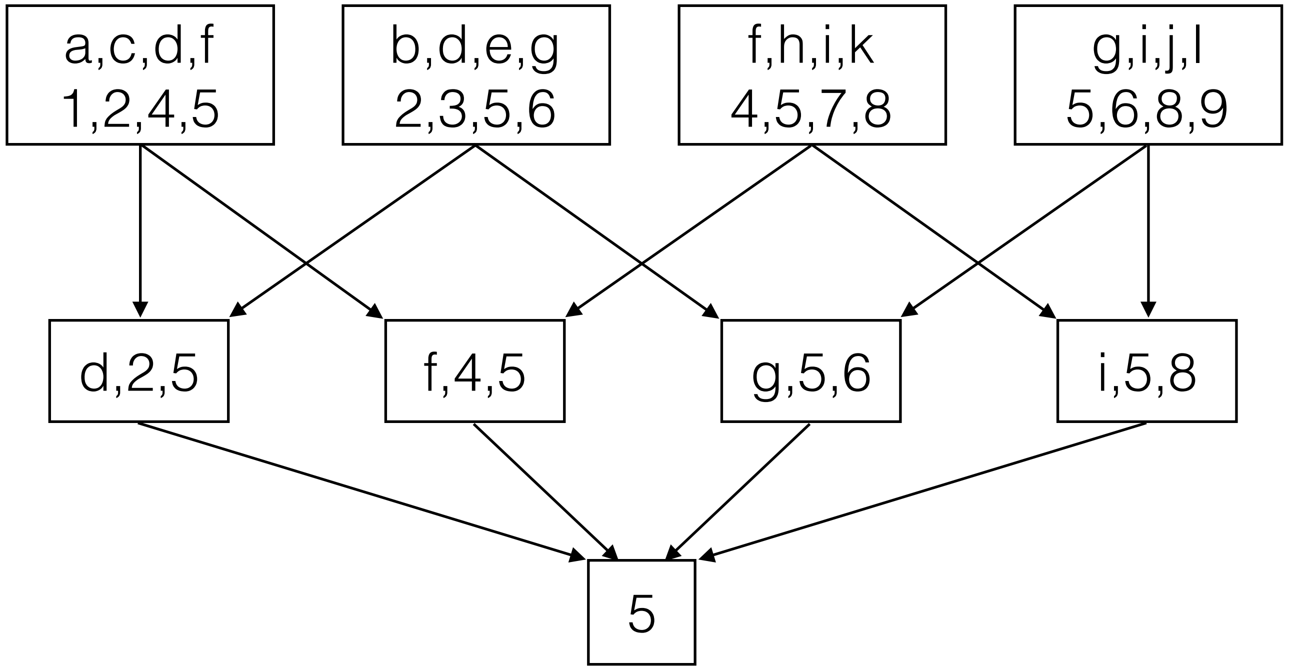




The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$




a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

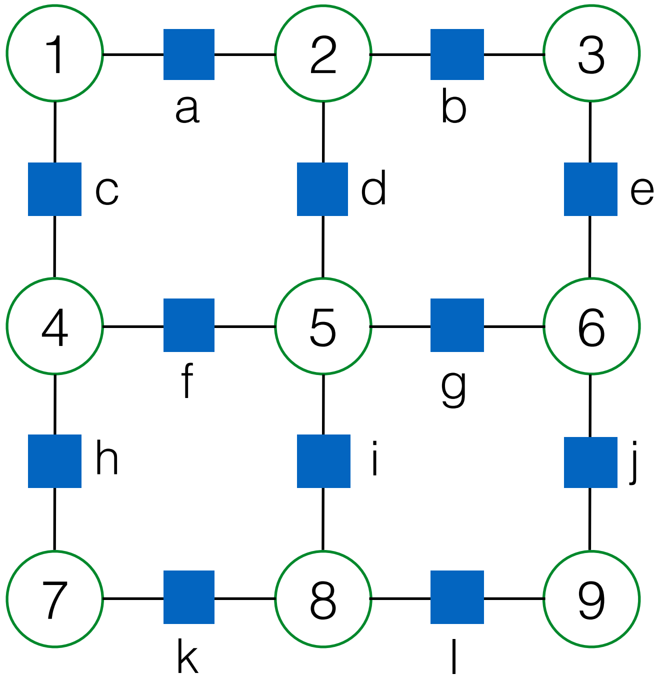
g,5,6

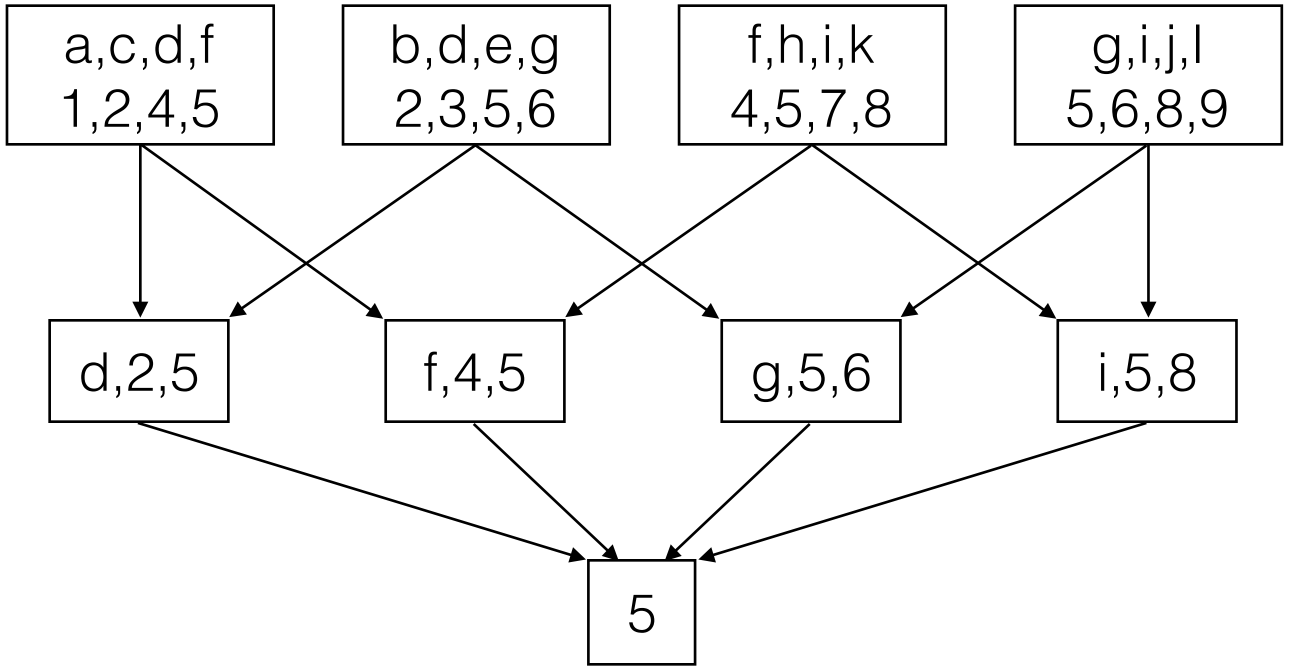
i,5,8

5

The Region Graph Method

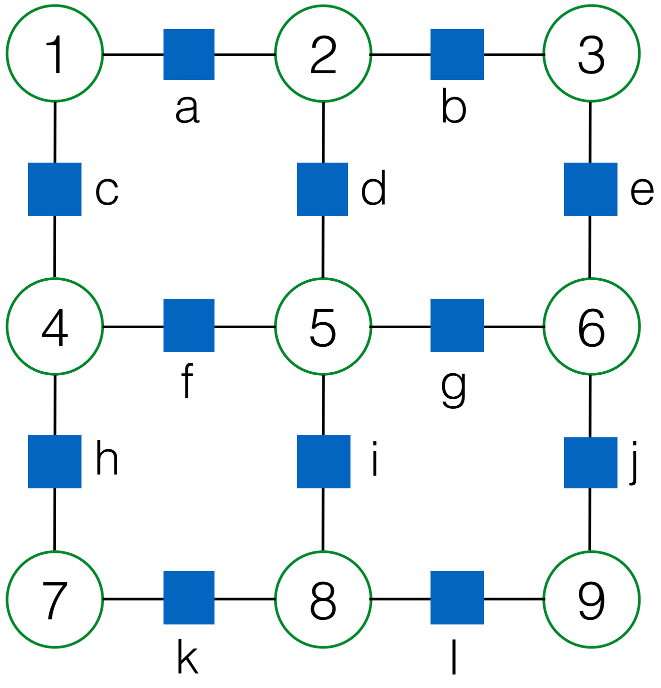
- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

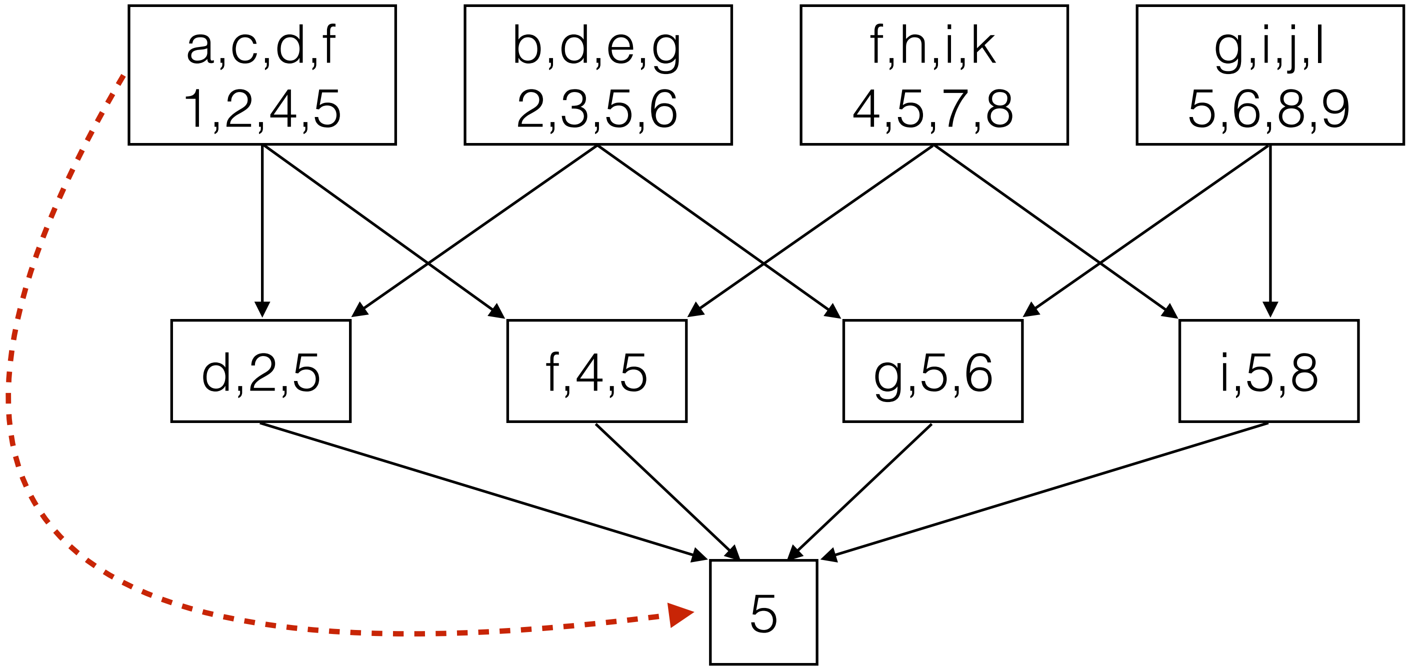




The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





a,c,d,f
1,2,4,5

b,d,e,g
2,3,5,6

f,h,i,k
4,5,7,8

g,i,j,l
5,6,8,9

d,2,5

f,4,5

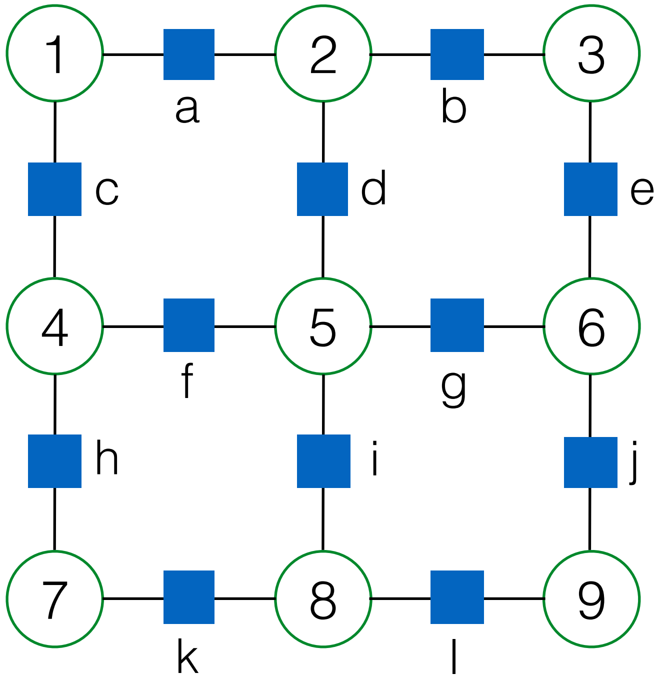
g,5,6

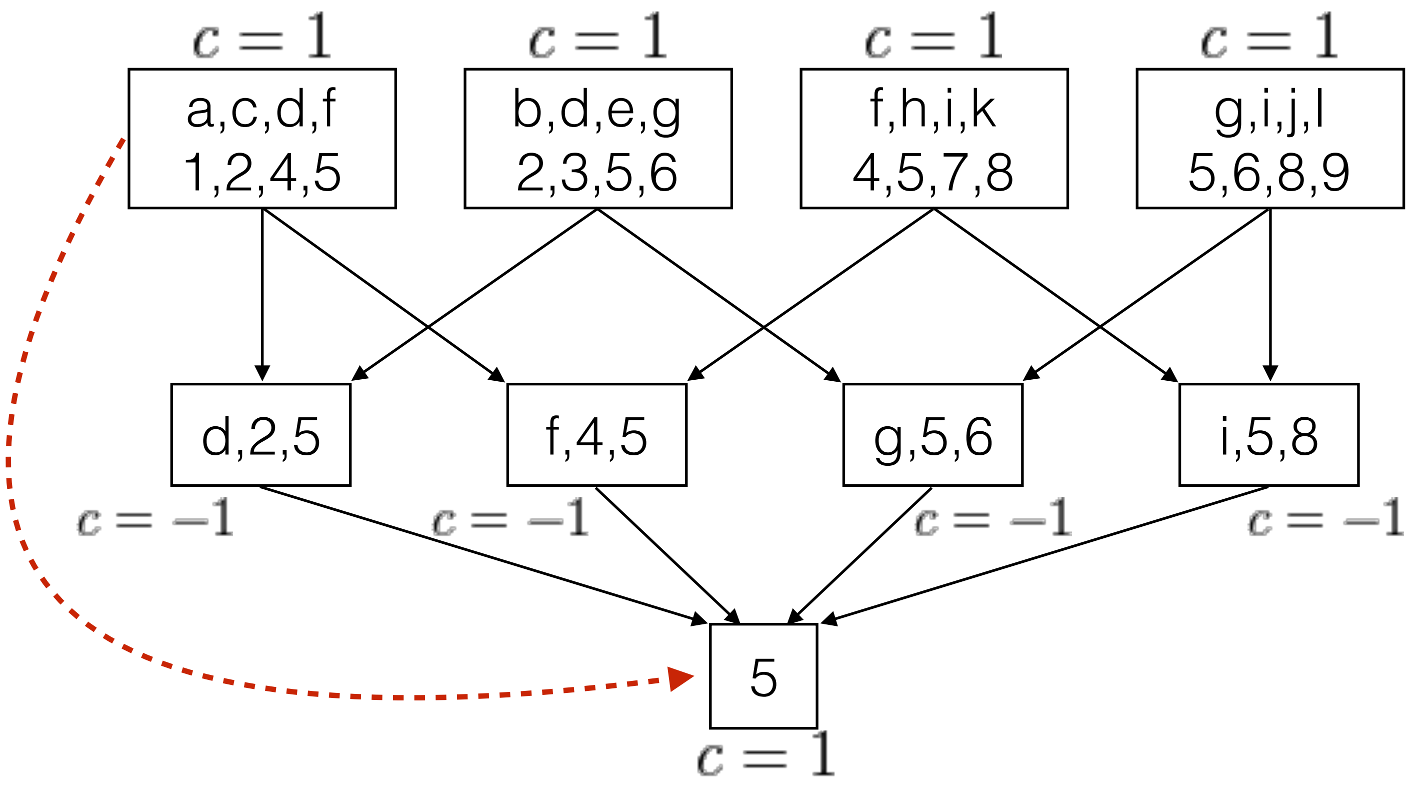
i,5,8

5

The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$

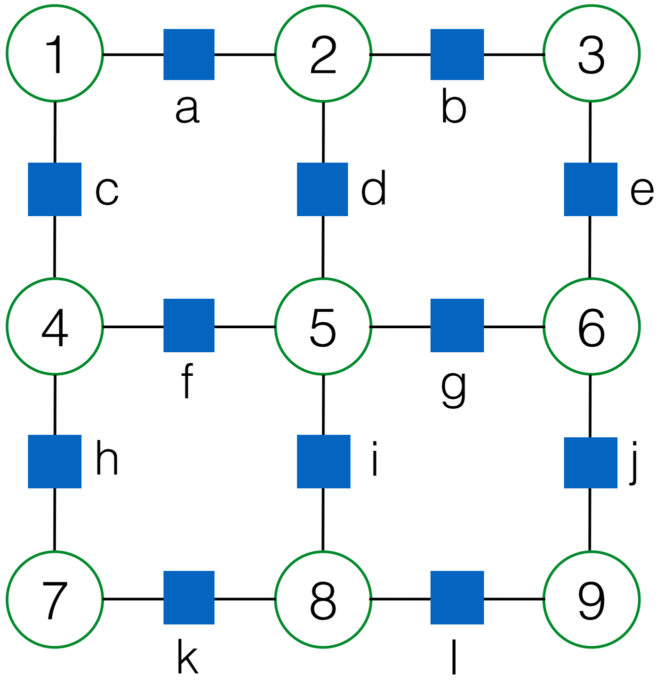


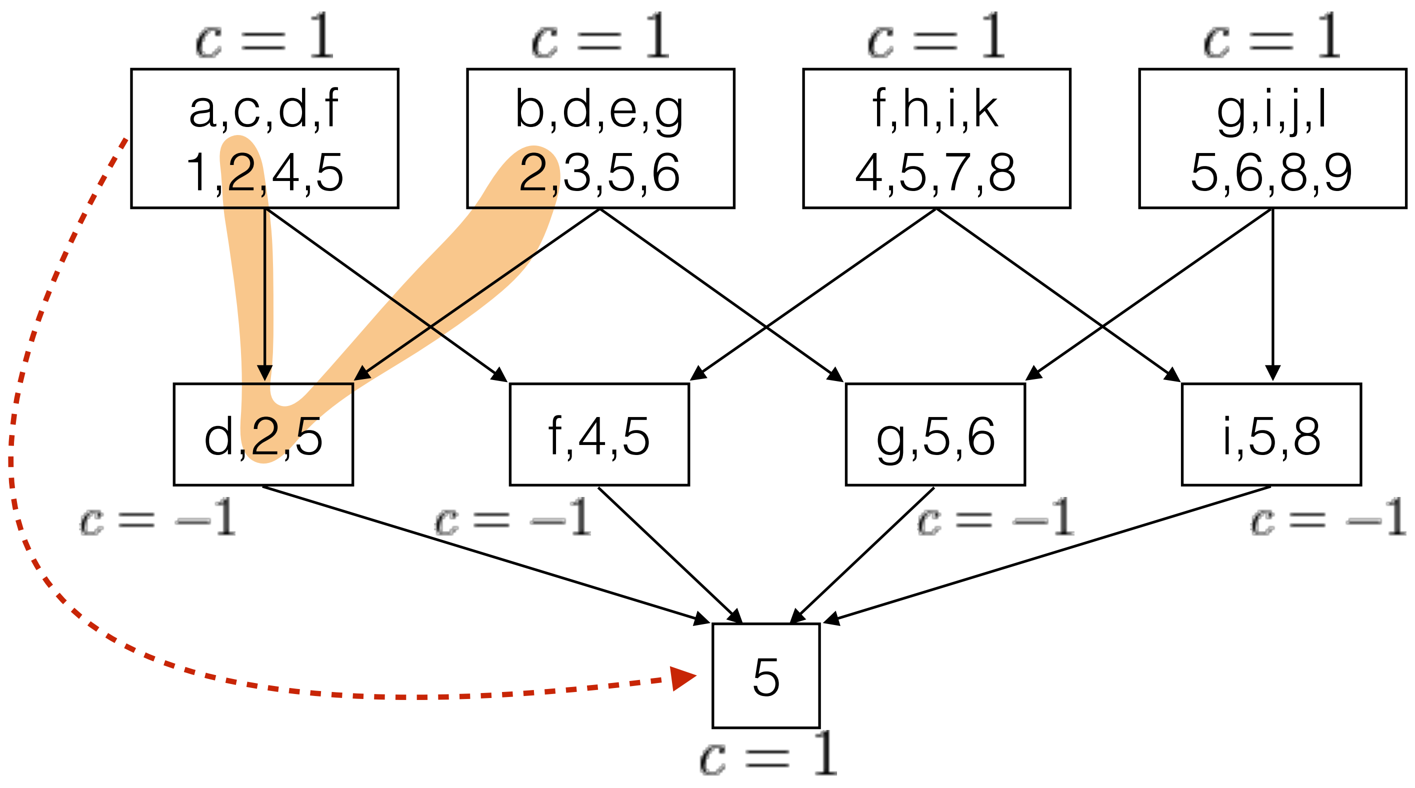


- Counting numbers: $c_v = 1 - \sum_{u \in \mathcal{A}(v)} c_u, \quad \forall v \in \mathcal{G}_{\mathrm{RG}}$ —> a valid approximation!

The Region Graph Method

- Definition: region graph $\mathcal{G}_{\mathrm{RG}} = (\mathcal{V}_{\mathrm{RG}}, \mathcal{E}_{\mathrm{RG}})$
  each vertex —> a region of the original factor graph $G = (V, A, E)$





- Counting numbers: $c_v = 1 - \sum_{u \in \mathcal{A}(v)} c_u, \quad \forall v \in \mathcal{G}_{\mathrm{RG}}$ —> a valid approximation!

- $\forall \alpha \in V \cup A \longrightarrow \mathcal{G}_{\mathrm{RG}}(\alpha)$ is a connected graph!

The Region Graph Method

(The Region-Based Approximation)

- The region-based (Gibbs) free energy approximation

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R \in \mathcal{R}} c_R F_R(b_R)$$

- Approximate free energy optimization problem:

$$\min_{\{b_R\}} F_{\mathcal{R}}(\{b_R\})$$

$$\text{s.t.} \quad \sum_{\boldsymbol{x}_P \setminus \boldsymbol{x}_C} b_P(\boldsymbol{x}_P) = b_C(\boldsymbol{x}_C)$$

$$\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) = 1$$

$$0 \leq b_R(\boldsymbol{x}_R) \leq 1$$

The Region Graph Method

(The Region-Based Approximation)

- The region-based (Gibbs) free energy approximation

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R \in \mathcal{R}} c_R F_R(b_R)$$

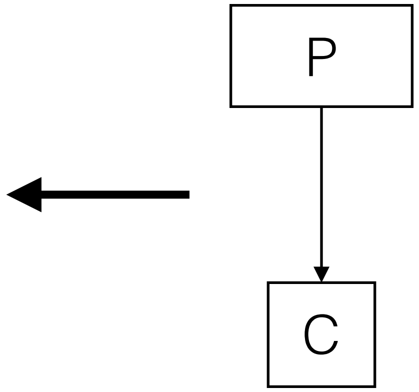
- Approximate free energy optimization problem:

$$\min_{\{b_R\}} F_{\mathcal{R}}(\{b_R\})$$

$$\text{s.t.} \quad \sum_{\boldsymbol{x}_P \setminus \boldsymbol{x}_C} b_P(\boldsymbol{x}_P) = b_C(\boldsymbol{x}_C)$$

$$\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) = 1$$

$$0 \leq b_R(\boldsymbol{x}_R) \leq 1$$



The Region Graph Method

(The Region-Based Approximation)

- The region-based (Gibbs) free energy approximation

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R \in \mathcal{R}} c_R F_R(b_R)$$

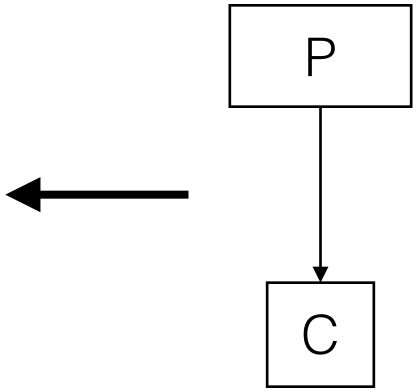
- Approximate free energy optimization problem:

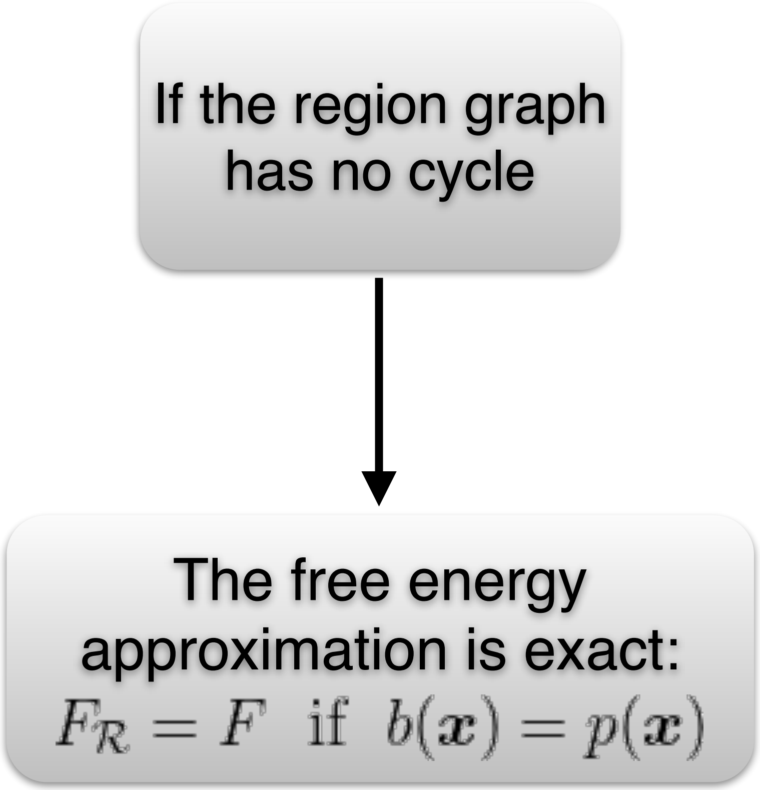
$$\min_{\{b_R\}} F_{\mathcal{R}}(\{b_R\})$$

$$\text{s.t.} \quad \sum_{\boldsymbol{x}_P \setminus \boldsymbol{x}_C} b_P(\boldsymbol{x}_P) = b_C(\boldsymbol{x}_C)$$

$$\sum_{\boldsymbol{x}_R} b_R(\boldsymbol{x}_R) = 1$$

$$0 \le b_R(\boldsymbol{x}_R) \le 1$$



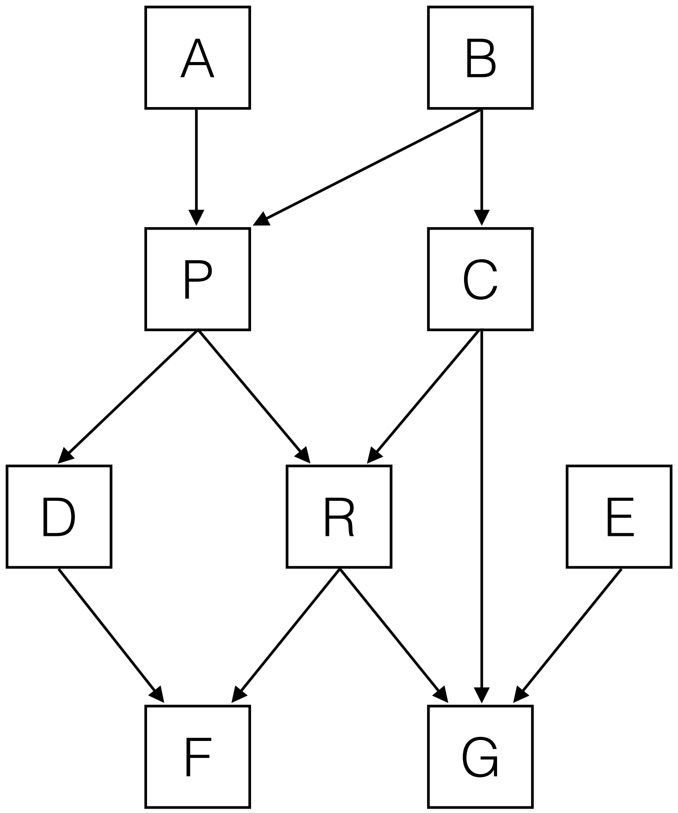


If the region graph has no cycle

The free energy approximation is exact: $F_{\mathcal{R}} = F \;\; \text{if} \;\; b(\boldsymbol{x}) = p(\boldsymbol{x})$

Generalized Belief Propagation
(The Parent to Child Algorithm)

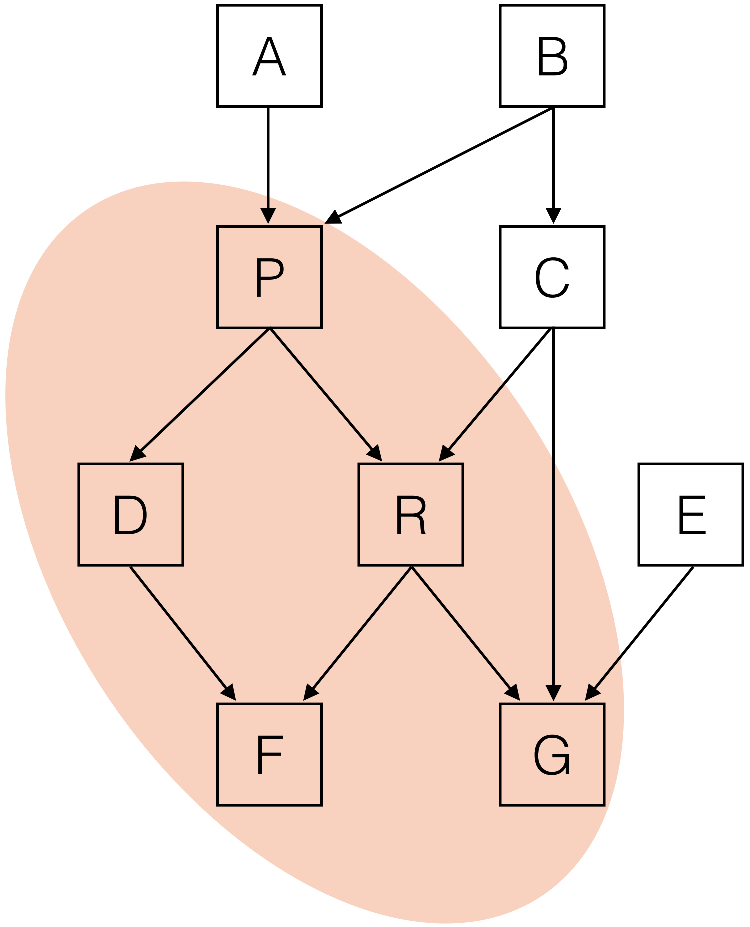
- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$



Generalized Belief Propagation

(The Parent to Child Algorithm)

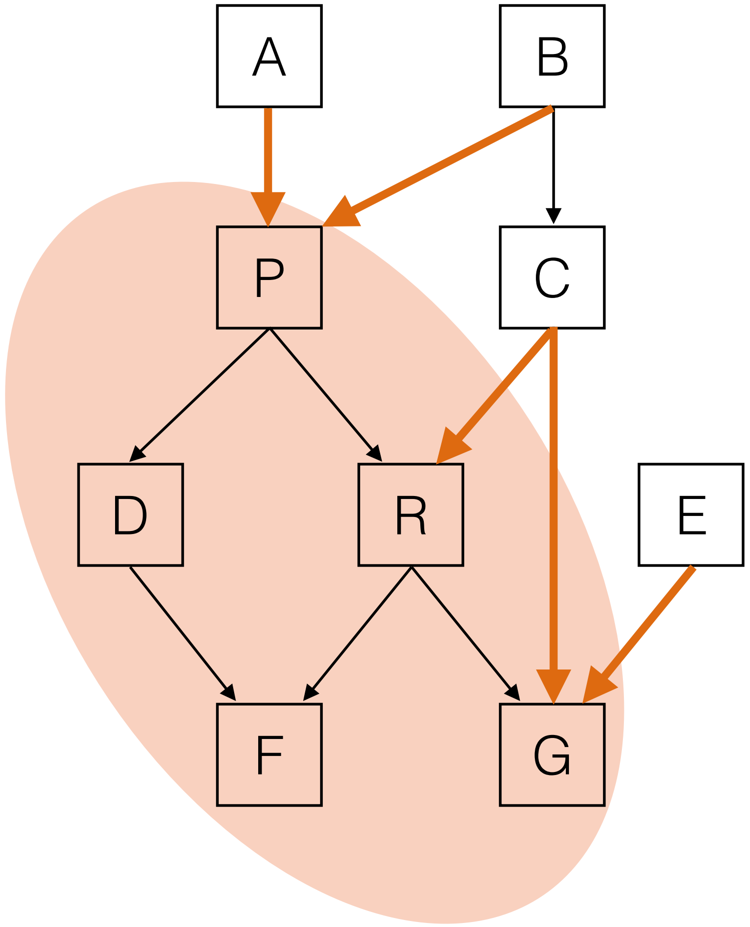
- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$



Generalized Belief Propagation

(The Parent to Child Algorithm)

- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$

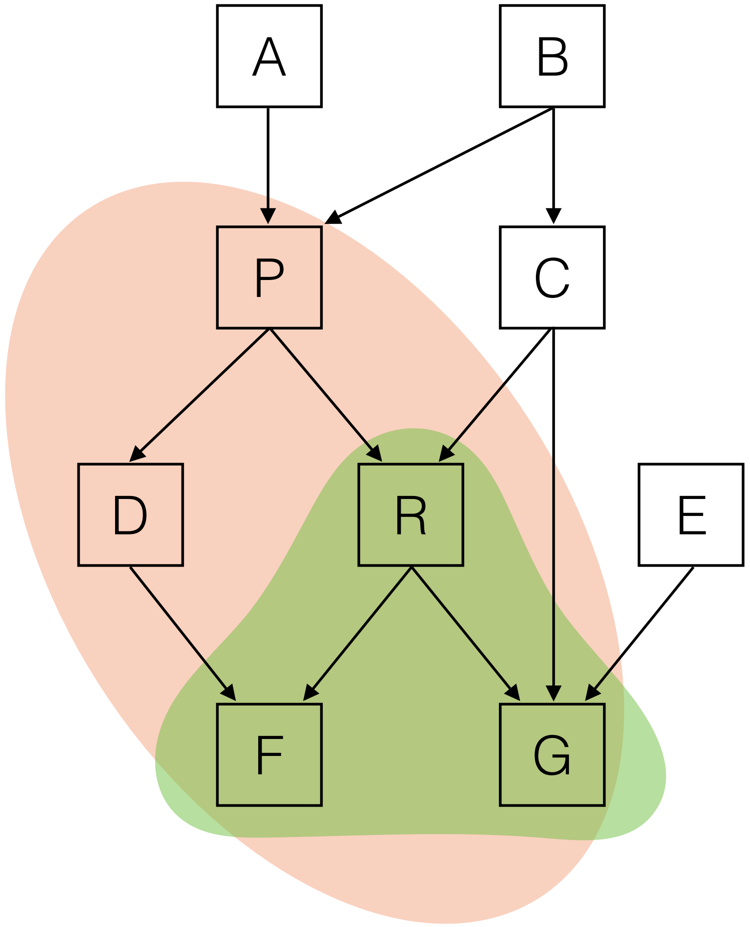


$$b_P \propto (m_{A \to P}\, m_{B \to P})\, (m_{C \to R}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_P} f_a(\boldsymbol{x}_a)$$

Generalized Belief Propagation

(The Parent to Child Algorithm)

- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$

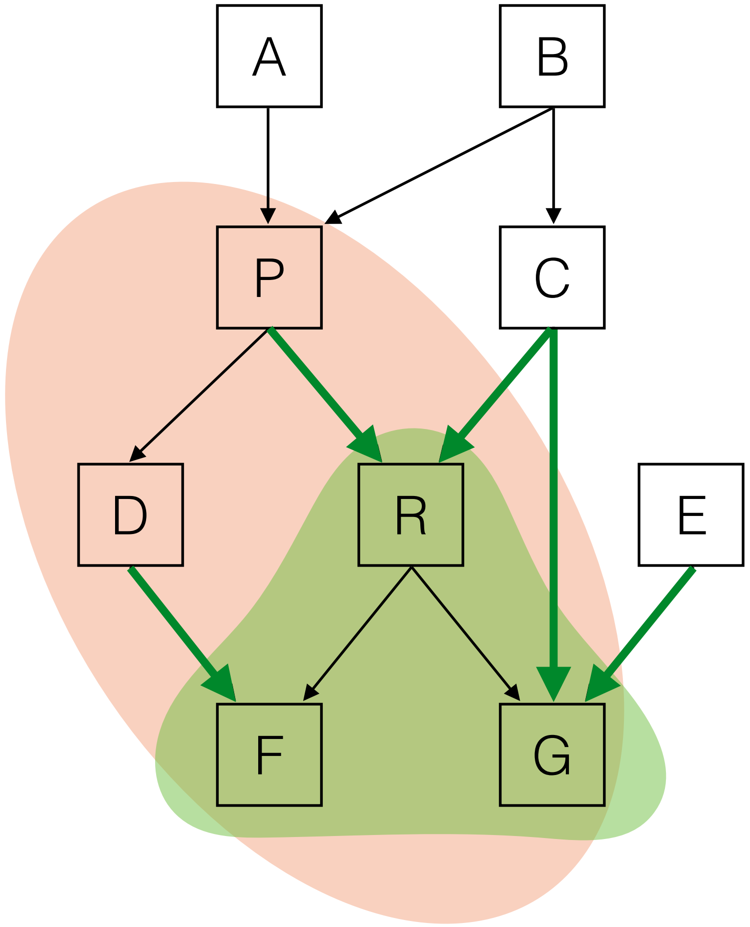


$$b_P \propto (m_{A \to P}\, m_{B \to P})\,(m_{C \to R}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_P} f_a(\boldsymbol{x}_a)$$

Generalized Belief Propagation

(The Parent to Child Algorithm)

- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$



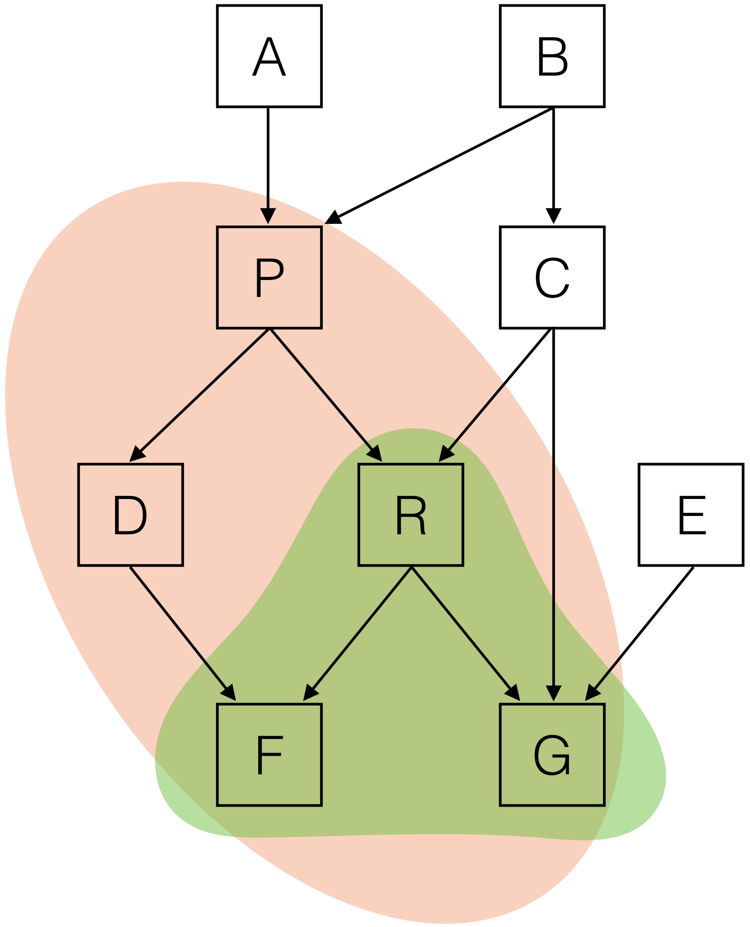
$$b_P \propto (m_{A \to P}\, m_{B \to P})\,(m_{C \to R}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_P} f_a(\boldsymbol{x}_a)$$

$$b_R \propto (m_{P \to R}\, m_{C \to R})\,(m_{D \to F}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_R} f_a(\boldsymbol{x}_a)$$

Generalized Belief Propagation

(The Parent to Child Algorithm)

- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$



$$b_P \propto (m_{A \to P}\, m_{B \to P})\,(m_{C \to R}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_P} f_a(\boldsymbol{x}_a)$$

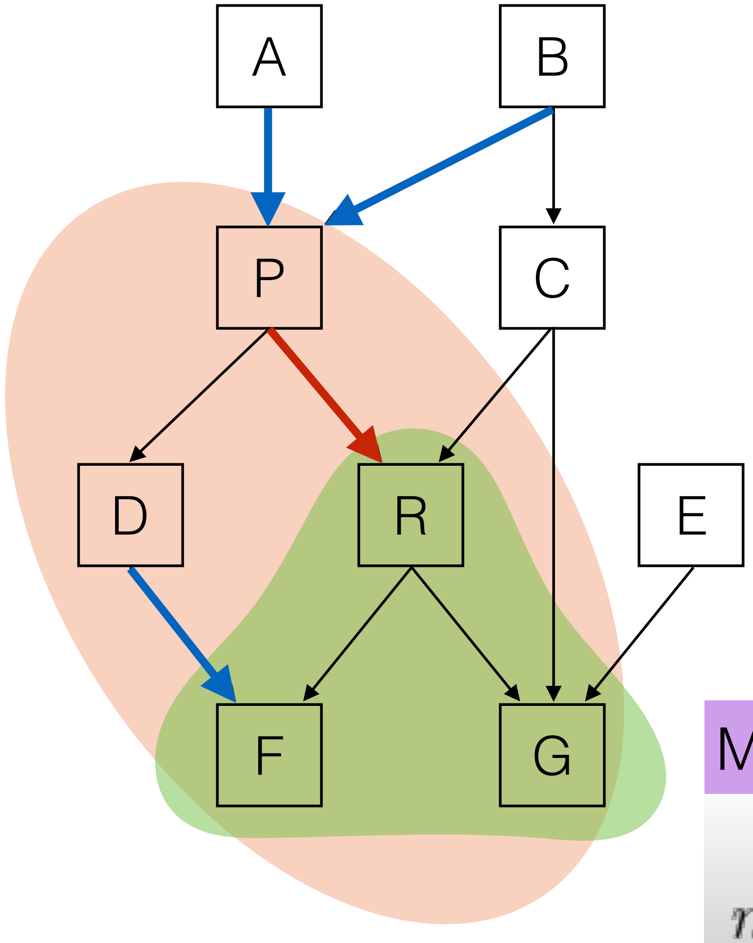
$$b_R \propto (m_{P \to R}\, m_{C \to R})\,(m_{D \to F}\, m_{C \to G}\, m_{E \to G}) \prod_{a \in A_R} f_a(\boldsymbol{x}_a)$$

$$b_R(\boldsymbol{x}_R) = \sum_{\boldsymbol{x}_P \setminus \boldsymbol{x}_R} b_P(\boldsymbol{x}_P)$$

Generalized Belief Propagation

(The Parent to Child Algorithm)

- We have only one kind of message $m_{P \to R}(\boldsymbol{x}_R)$



$$b_P \propto (m_{A\to P}\, m_{B\to P})\,(m_{C\to R}\, m_{C\to G}\, m_{E\to G}) \prod_{a \in A_P} f_a(\boldsymbol{x}_a)$$

$$b_R \propto (m_{P\to R}\, m_{C\to R})\,(m_{D\to F}\, m_{C\to G}\, m_{E\to G}) \prod_{a \in A_R} f_a(\boldsymbol{x}_a)$$

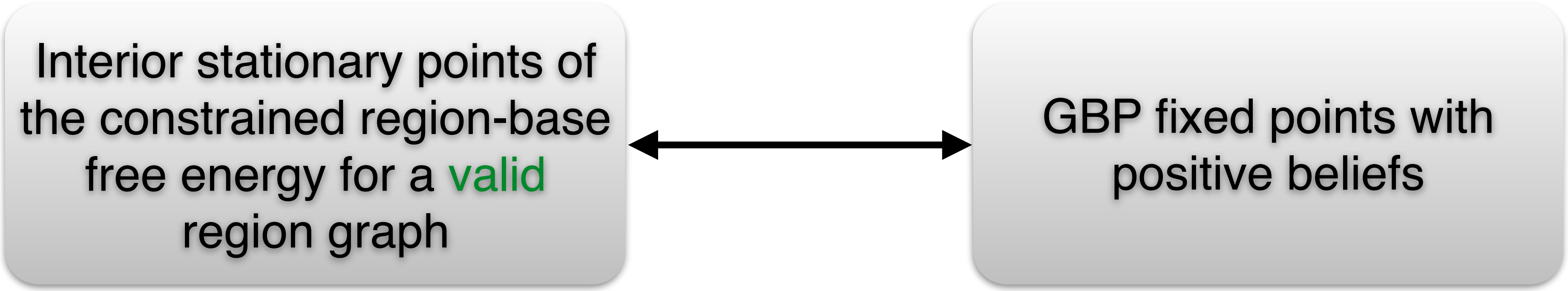
$$b_R(\boldsymbol{x}_R) = \sum_{\boldsymbol{x}_P \backslash \boldsymbol{x}_R} b_P(\boldsymbol{x}_P)$$

Message Update Rules

$$m_{P\to R}(\boldsymbol{x}_R) = \frac{\sum_{\boldsymbol{x}_P \backslash \boldsymbol{x}_R} m_{A\to P}(\boldsymbol{x}_P) m_{B\to P}(\boldsymbol{x}_P) \prod_{a \in A_P \backslash A_R} f_a(\boldsymbol{x}_a)}{m_{D\to F}(\boldsymbol{x}_F)}$$

Connection Between
Region Graph Method and GBP

- Theorem:



Interior stationary points of the constrained region-base free energy for a valid region graph

GBP fixed points with positive beliefs

- In contrast to Bethe approximation:
people started from the region-based approximation and using Lagrange method derived the GBP algorithm

Generalized Belief Propagation

- Generalized belief propagation has other variations:
  - Parent to child algorithm
  - Child to parent algorithm
  - two-way algorithm
- The BP algorithm is a special case of all the above algorithms if the regions are chosen according to Bethe approximation
- The GBP is more complex than BP but it provides more flexibility in terms of choosing the regions (i.e. how to approximate Gibbs free energy)

Generalized Belief Propagation for Estimating the Partition Function of the 2D Ising Model

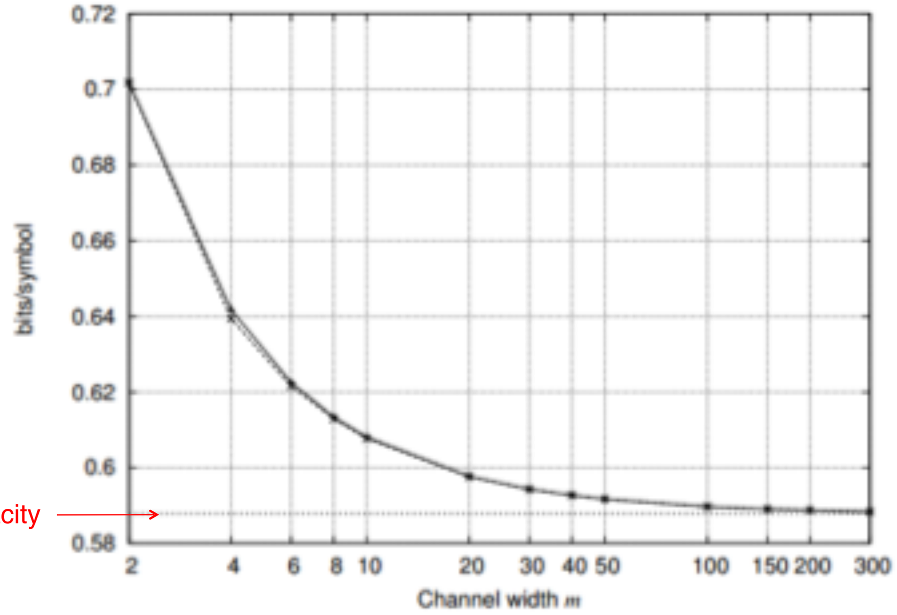
joint work with

Chun Lam Chan, Sidharth Jaggi, Navin Kashyap, and Pascal O. Vontobel

2D Ising Model

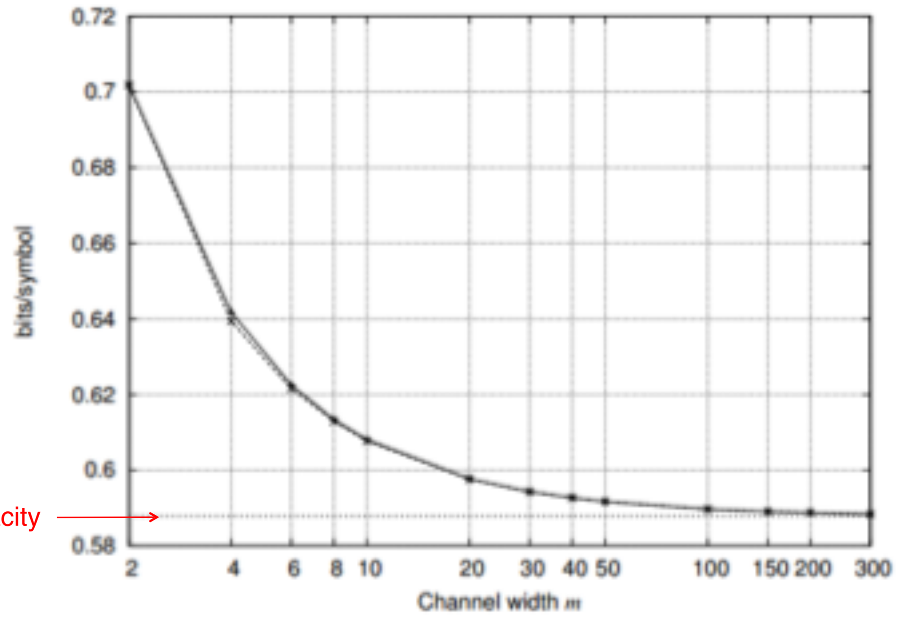
- Motivated by a 2D run-length limited (RLL) constraints problem
  - A symmetric (d, k) RLL constraint imposes (horizontally and vertically):
    - At least d zero symbols between two ones
    - At most k zero symbols between two ones
- Sabato, G. and Molkaraie observed that GBP can potentially outperform BP approximating capacity of an RLL problem

Capacity of 2D (1,∞)-RLL Constraint



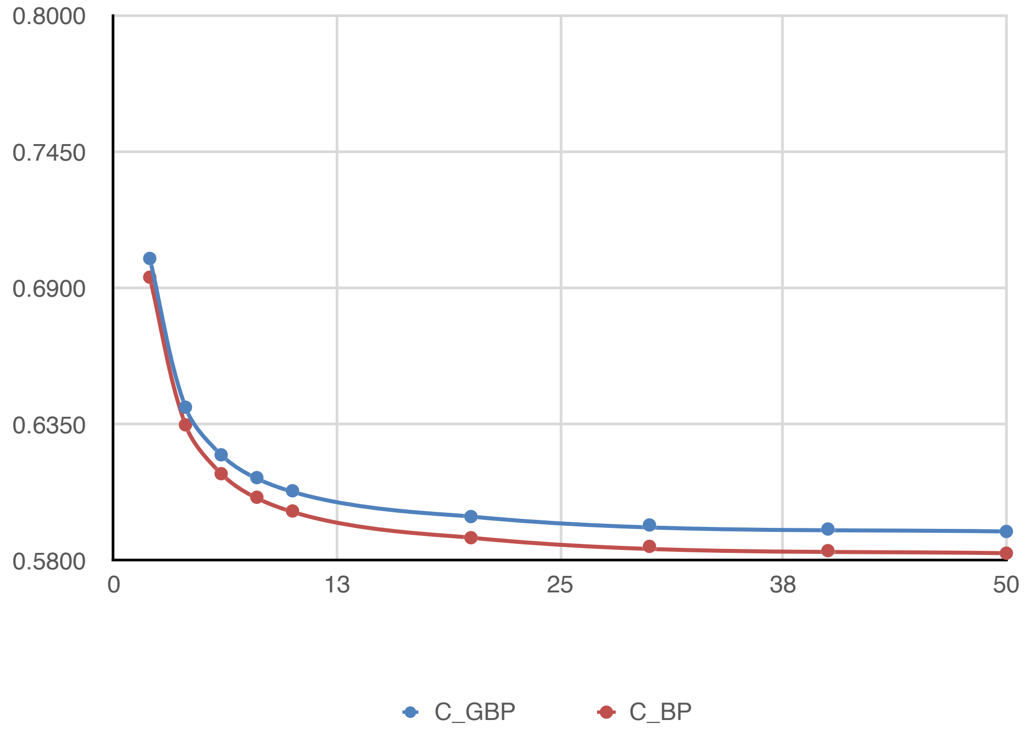
Shannon capacity

Capacity of 2D (1,∞)-RLL Constraint

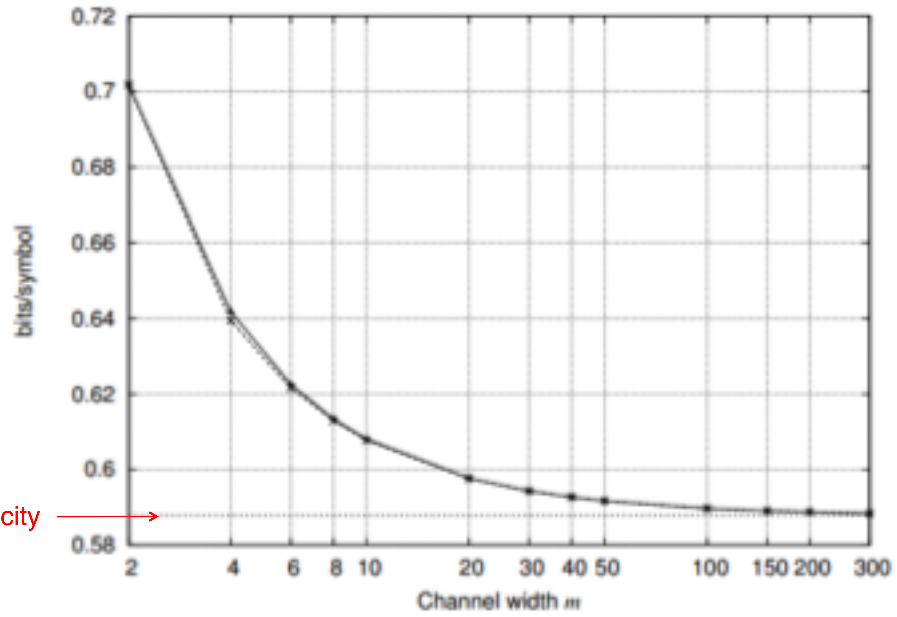


Shannon capacity →

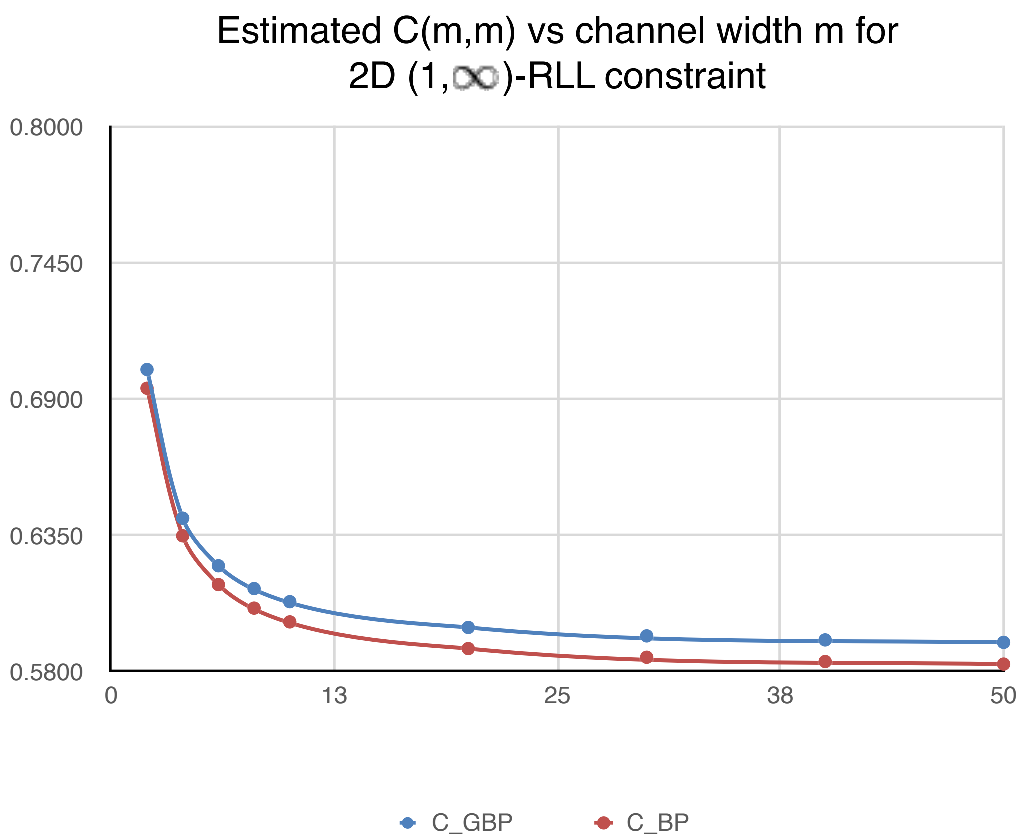
Estimated C(m,m) vs channel width m for 2D (1,∞)-RLL constraint



Capacity of 2D (1,∞)-RLL Constraint



Shannon capacity →

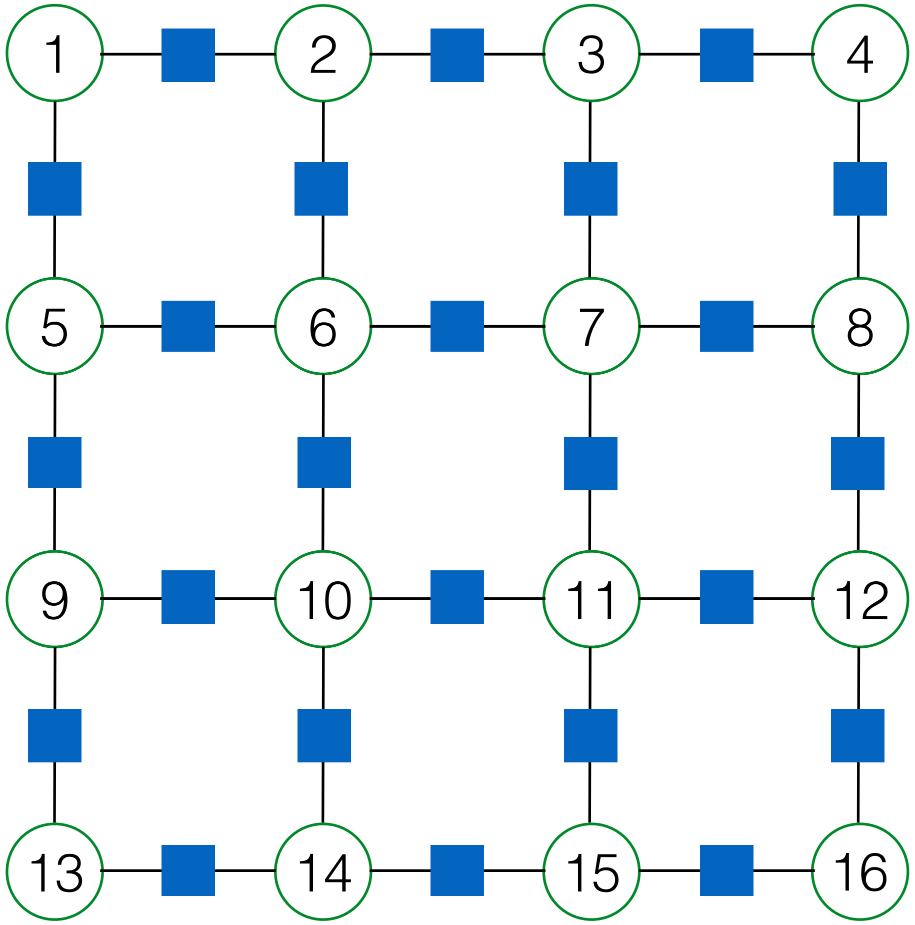

Estimated C(m,m) vs channel width m for 2D (1,∞)-RLL constraint

$$C(m,m) = \frac{\log_2 Z(m,m)}{m \times m}$$

2D Binary Ising Model

Binary variables: {0,1}

Homogenous pairwise function

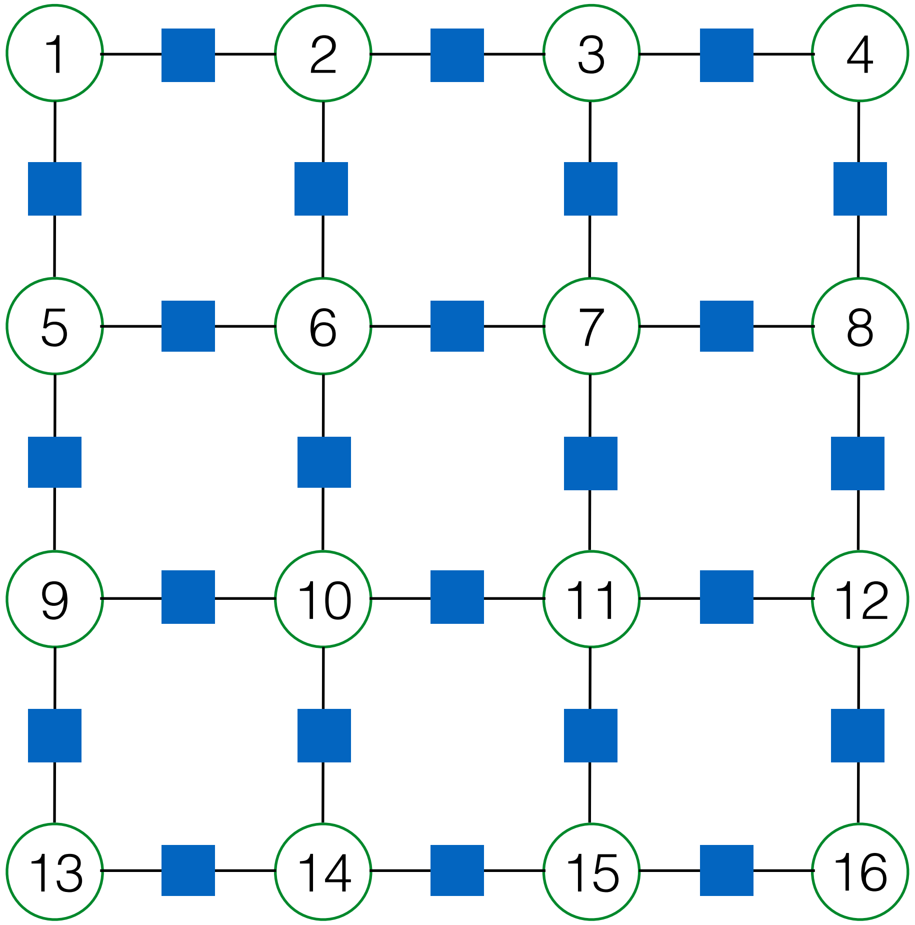


$$p(\boldsymbol{x}) = \frac{1}{Z}\prod_{a\in A} f_a(\boldsymbol{x}_a)$$

$$Z = \sum_{\boldsymbol{x}}\prod_{a\in A} f_a(\boldsymbol{x}_a)$$

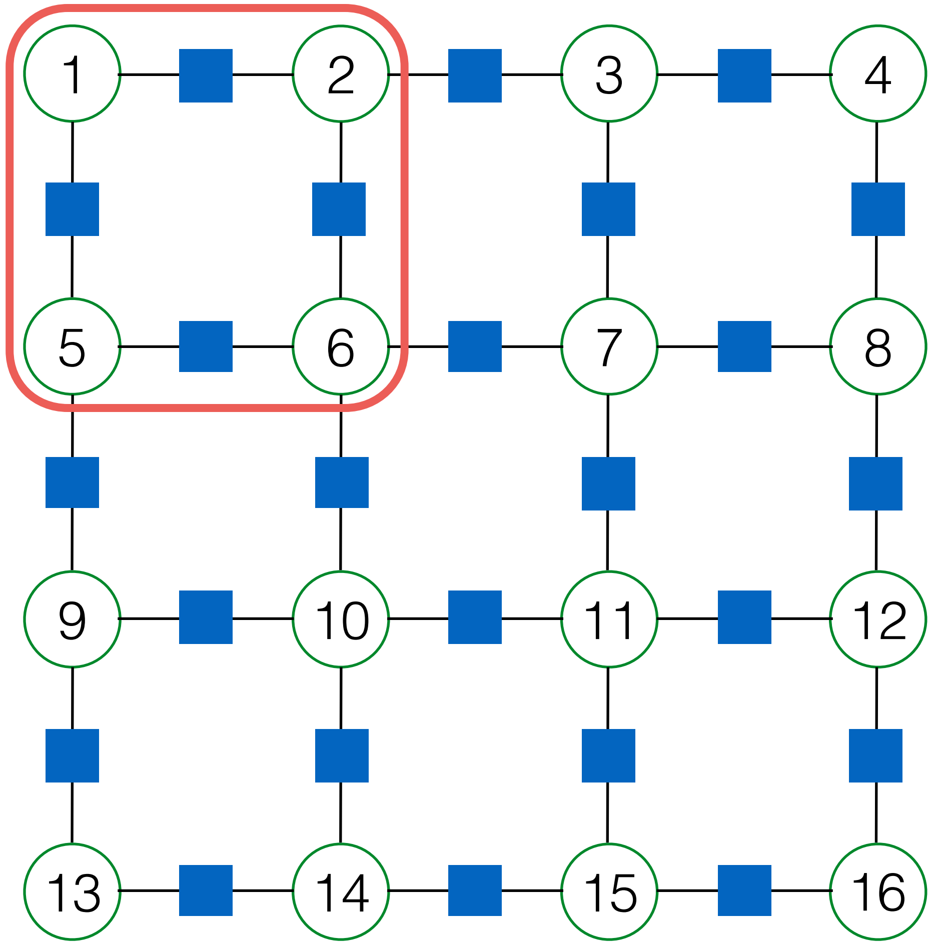
Region-Based Approximation

(The Choice of Regions)



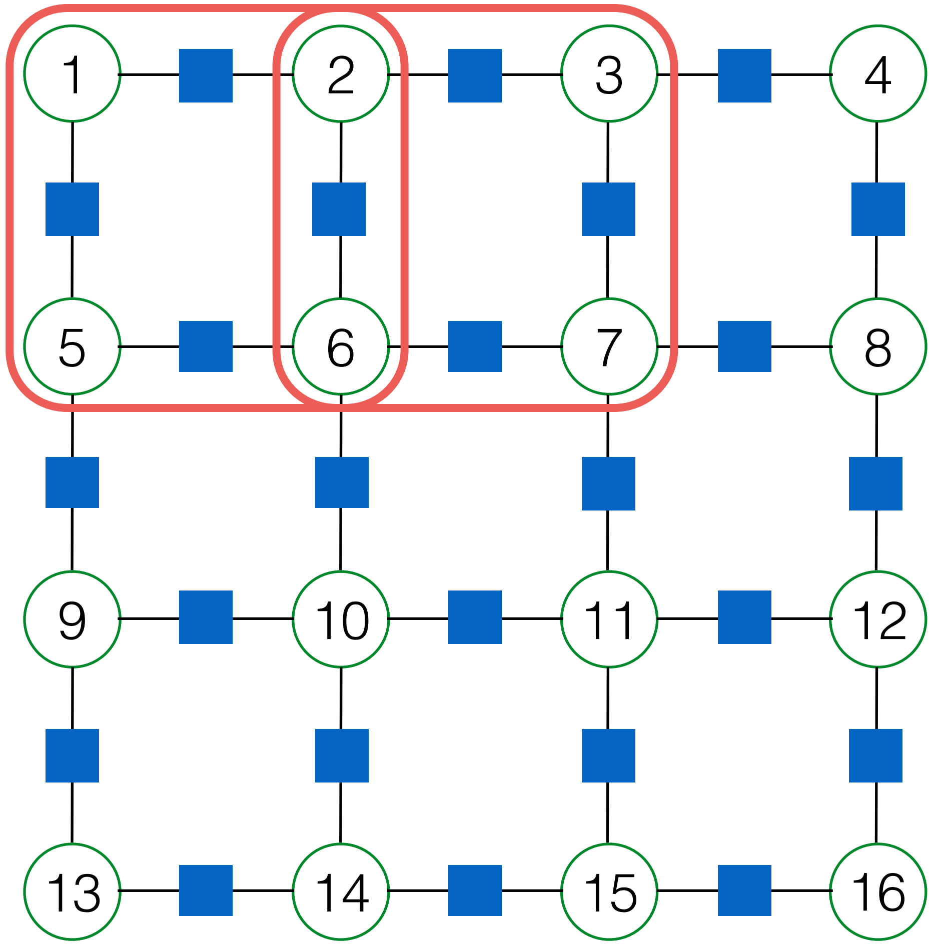
Region-Based Approximation

(The Choice of Regions)



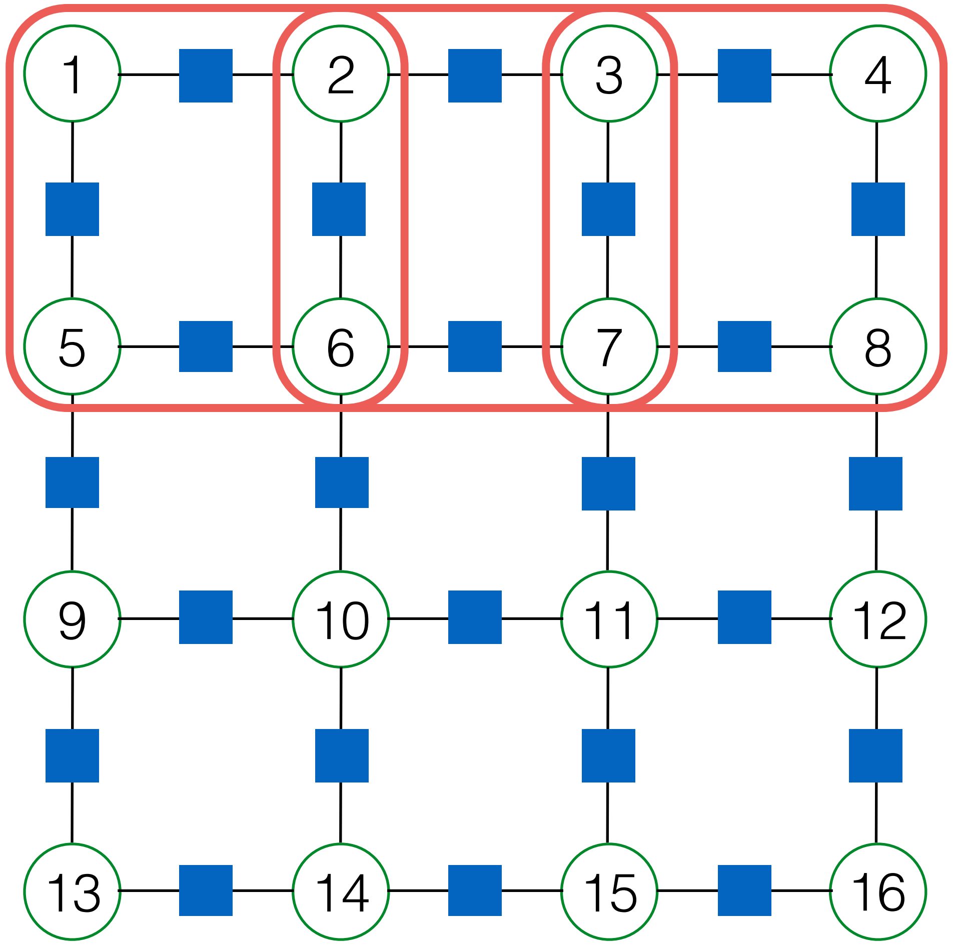
Region-Based Approximation

(The Choice of Regions)



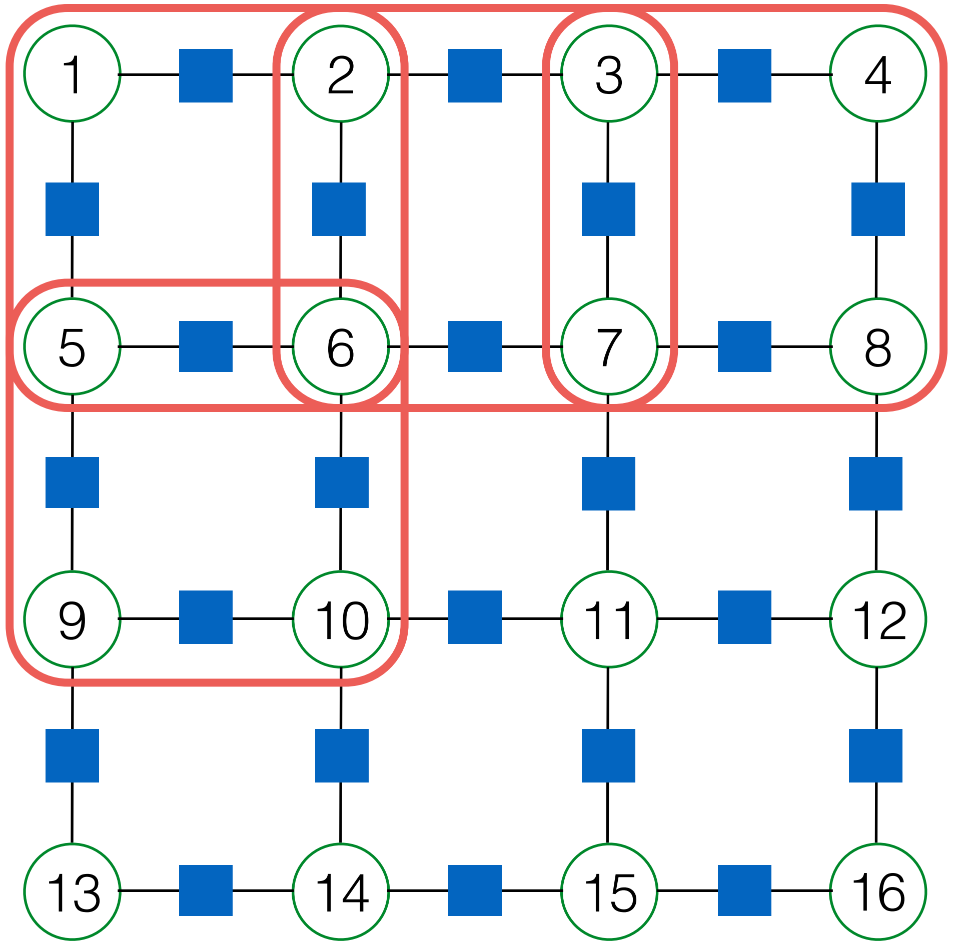
Region-Based Approximation

(The Choice of Regions)



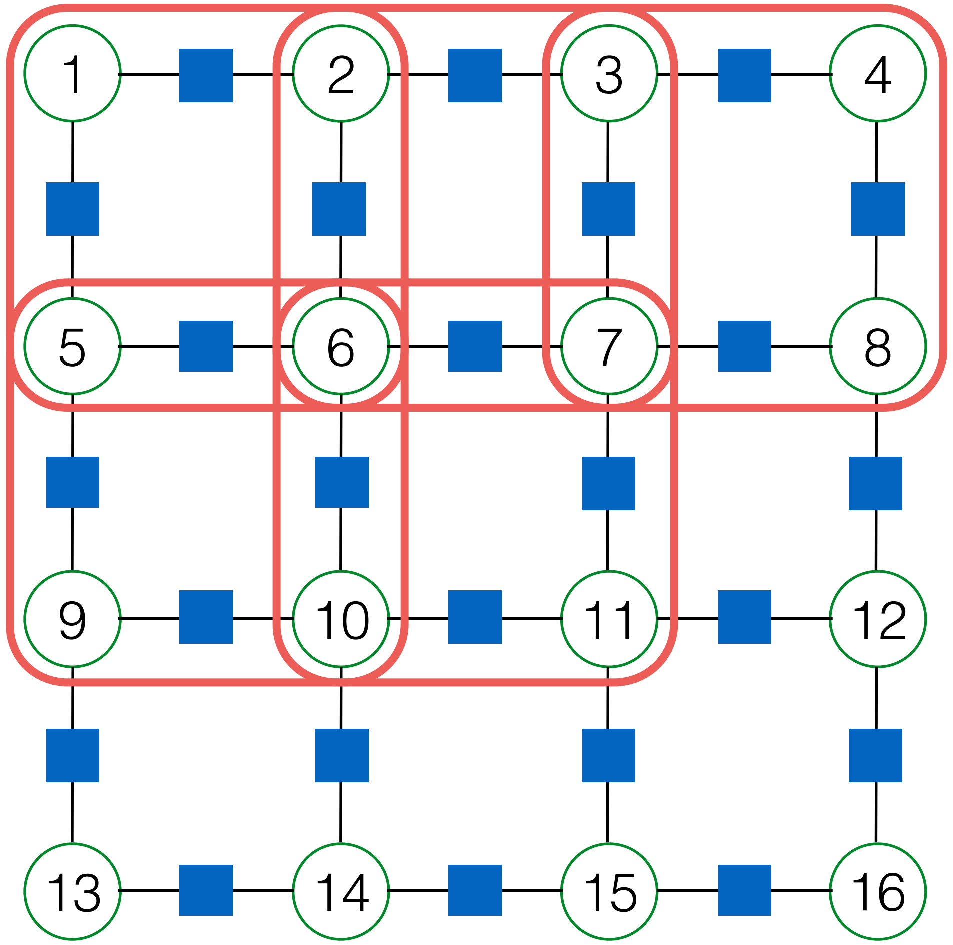
Region-Based Approximation

(The Choice of Regions)



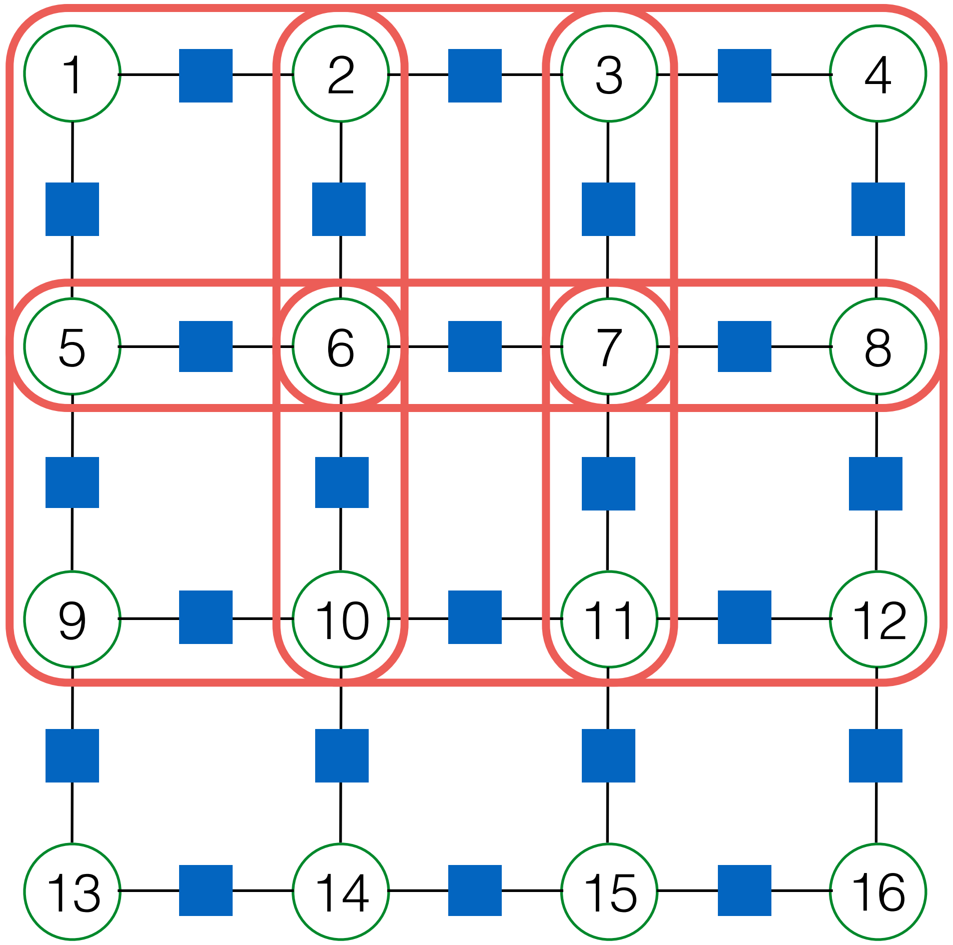
Region-Based Approximation

(The Choice of Regions)



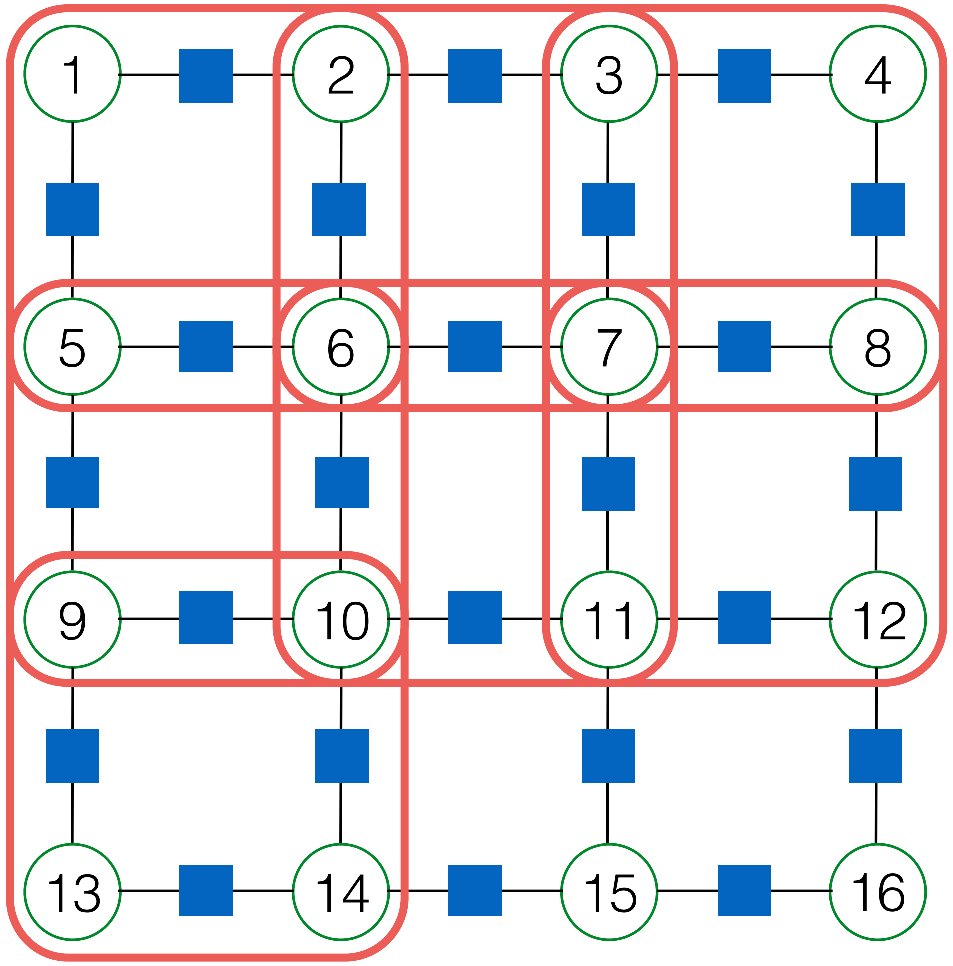
Region-Based Approximation

(The Choice of Regions)



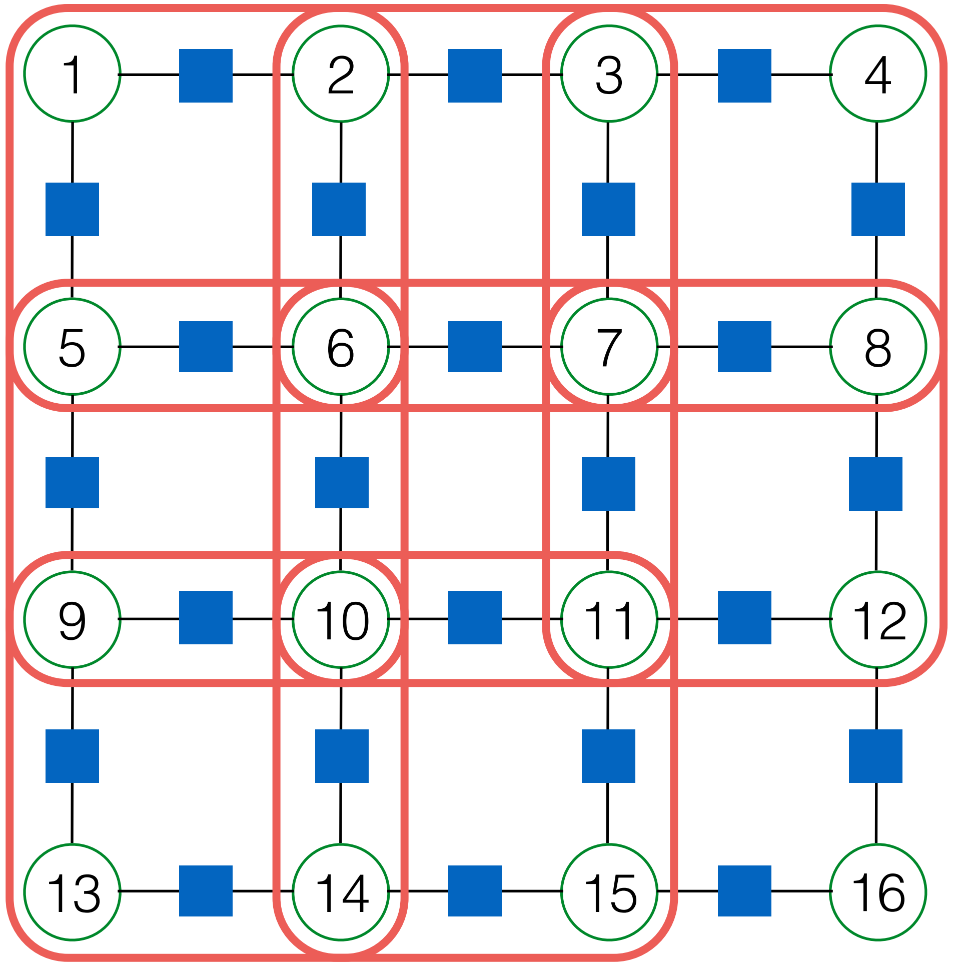
Region-Based Approximation

(The Choice of Regions)



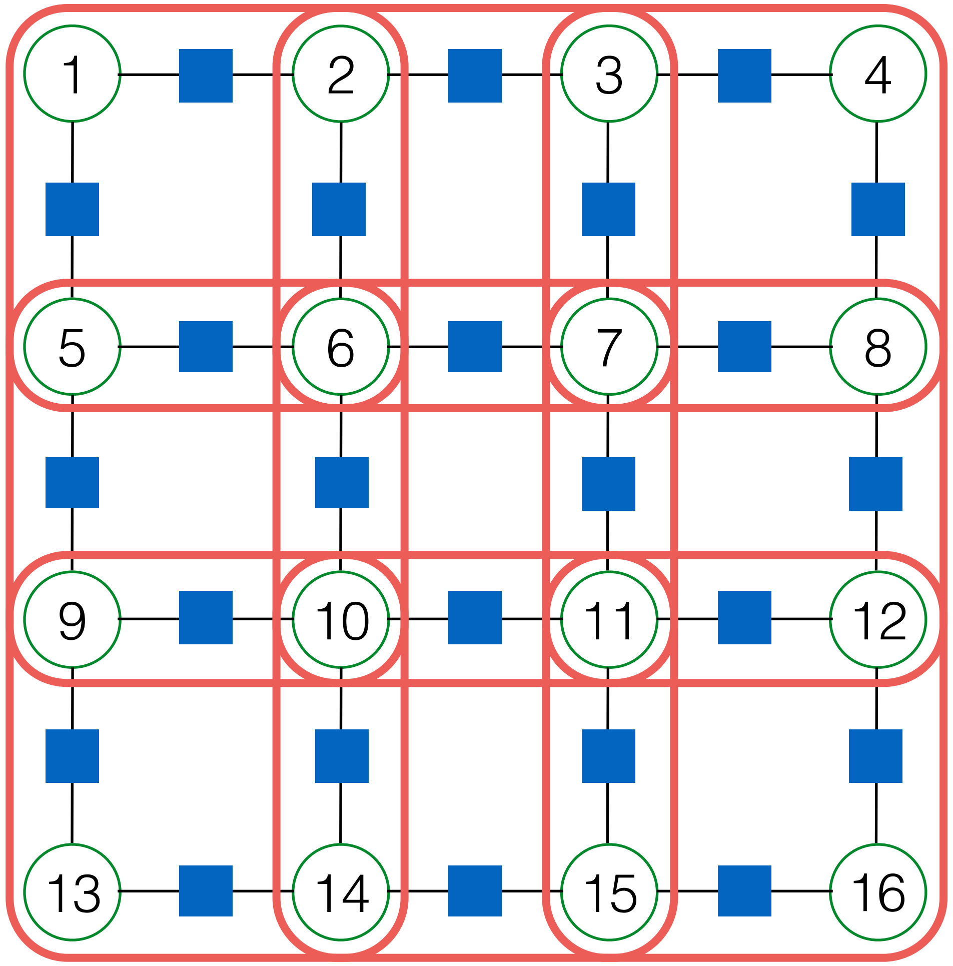
Region-Based Approximation

(The Choice of Regions)



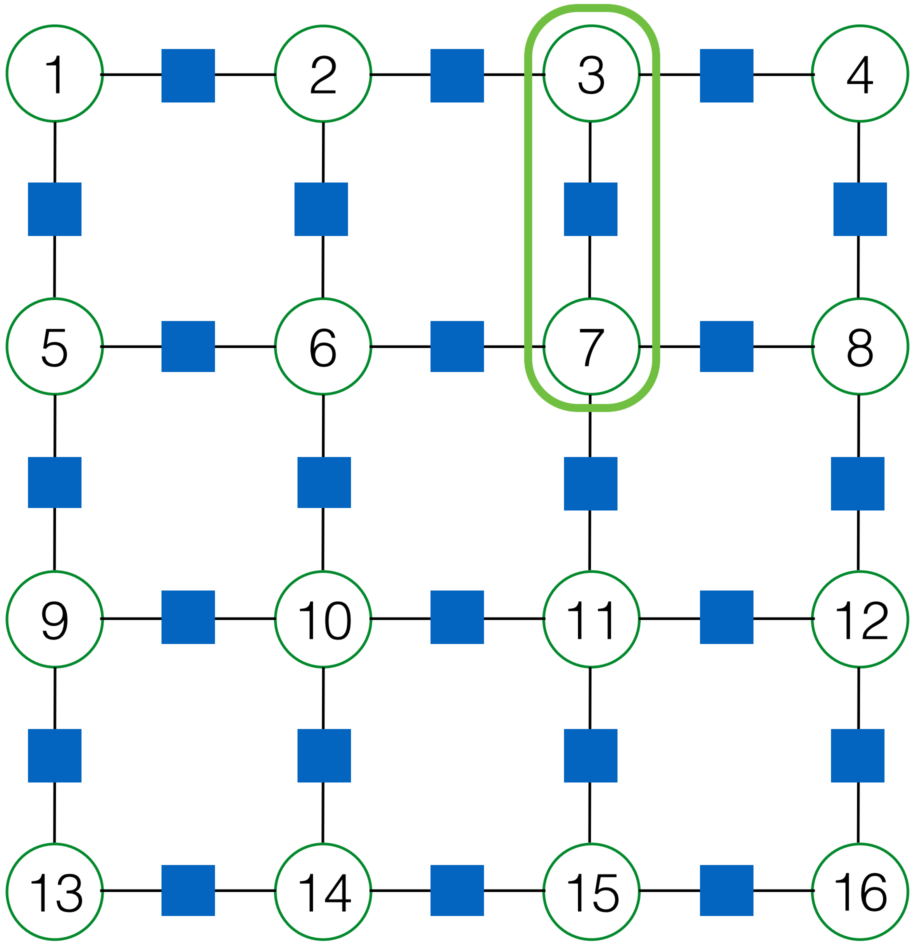
Region-Based Approximation

(The Choice of Regions)



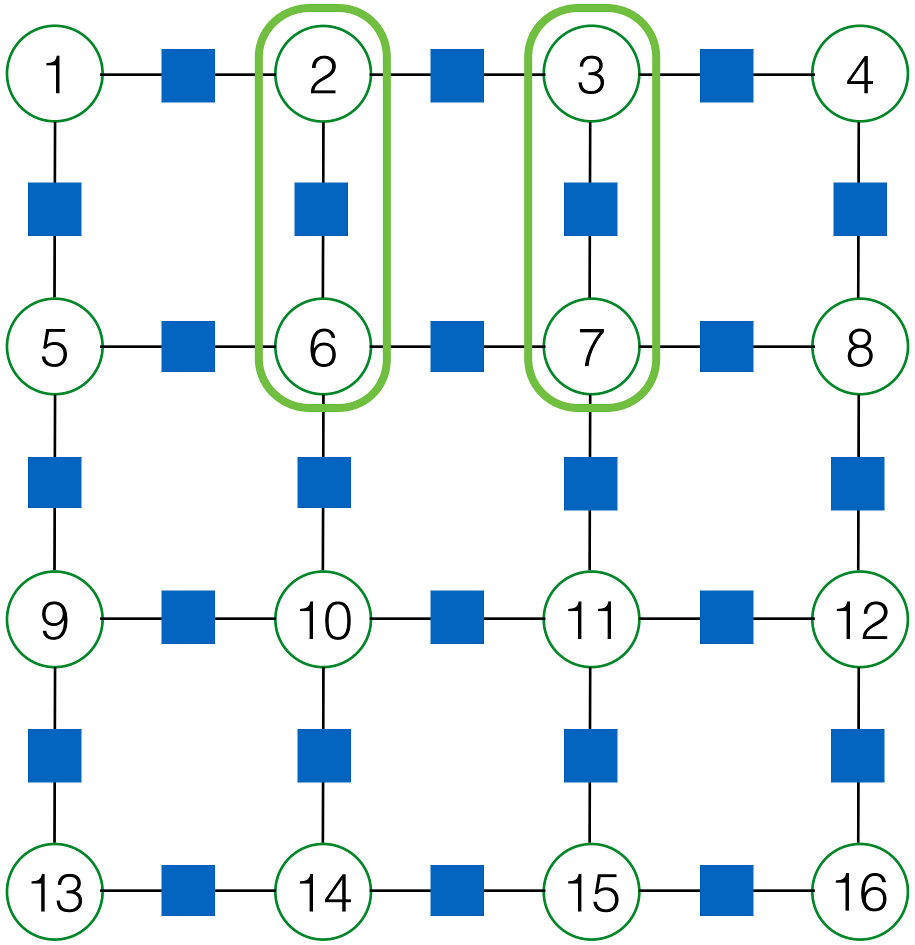
Region-Based Approximation

(The Choice of Regions)



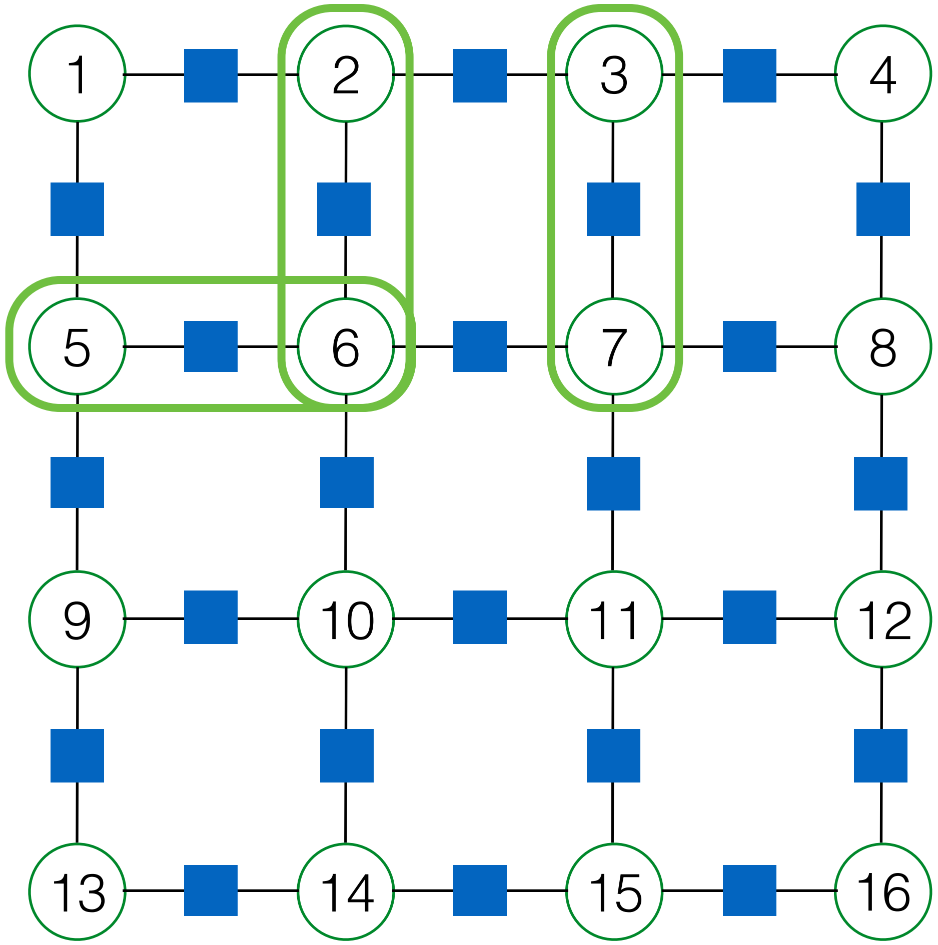
Region-Based Approximation

(The Choice of Regions)



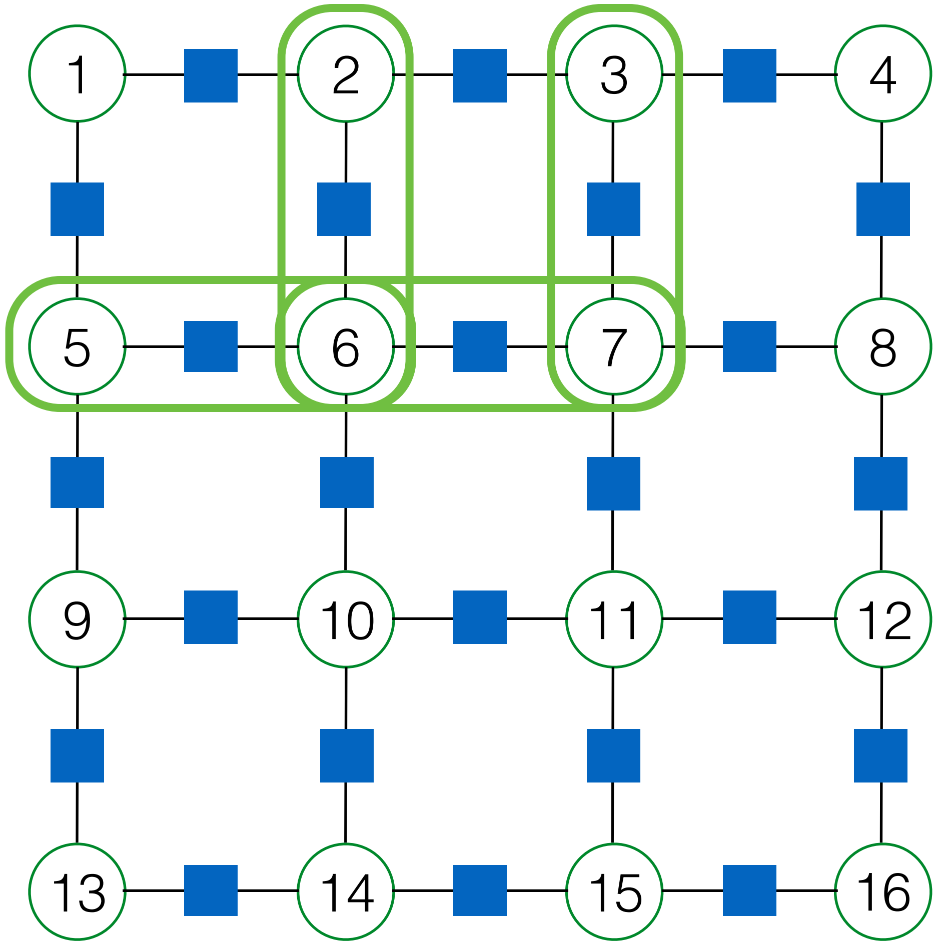
Region-Based Approximation

(The Choice of Regions)



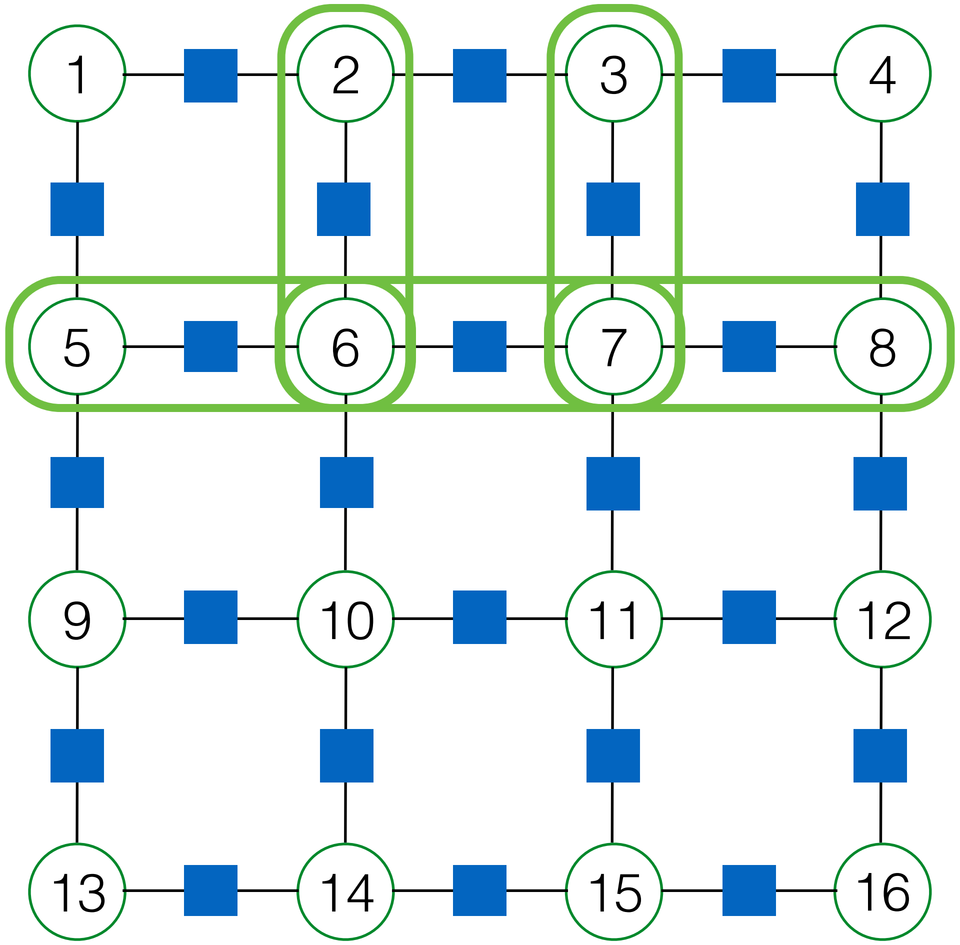
Region-Based Approximation

(The Choice of Regions)



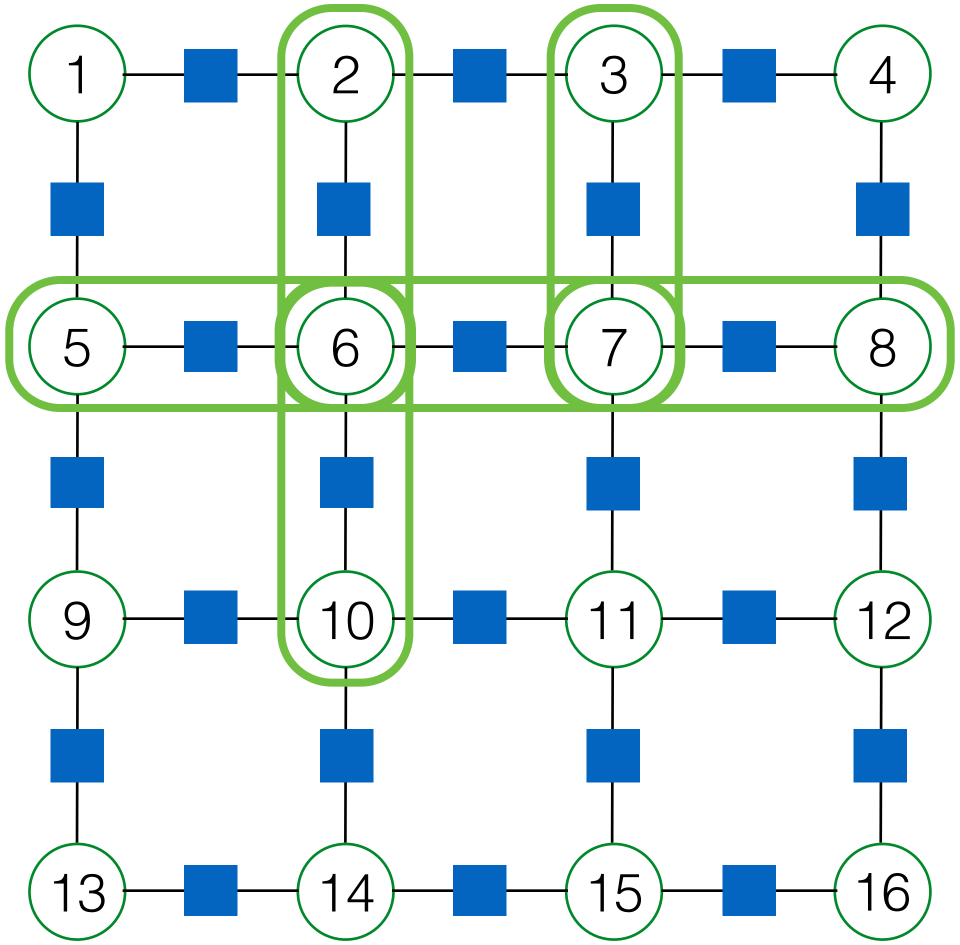
Region-Based Approximation

(The Choice of Regions)



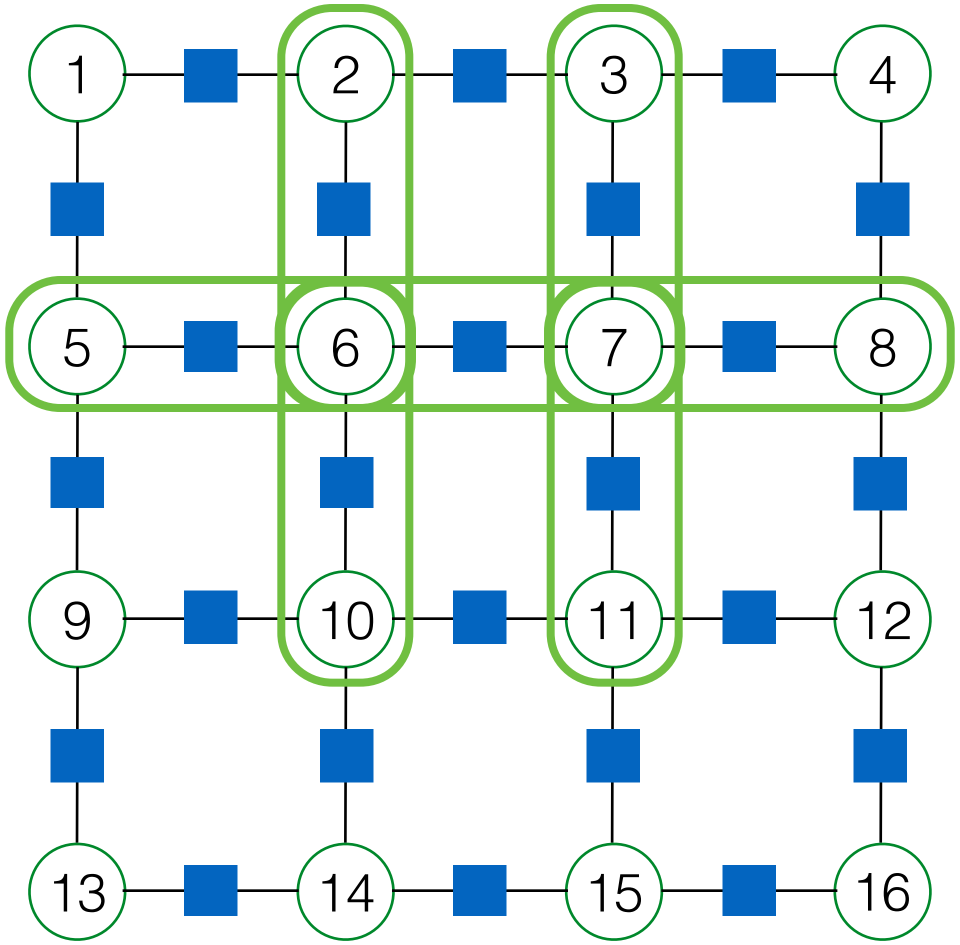
Region-Based Approximation

(The Choice of Regions)



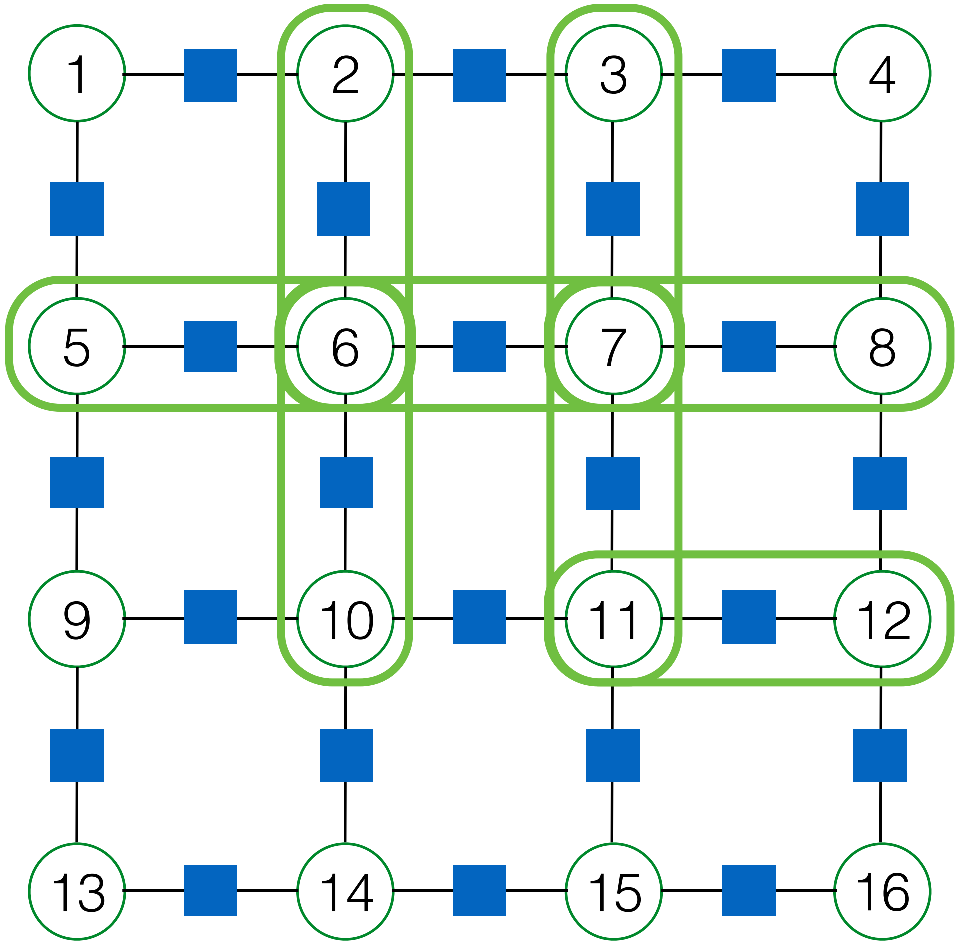
Region-Based Approximation

(The Choice of Regions)



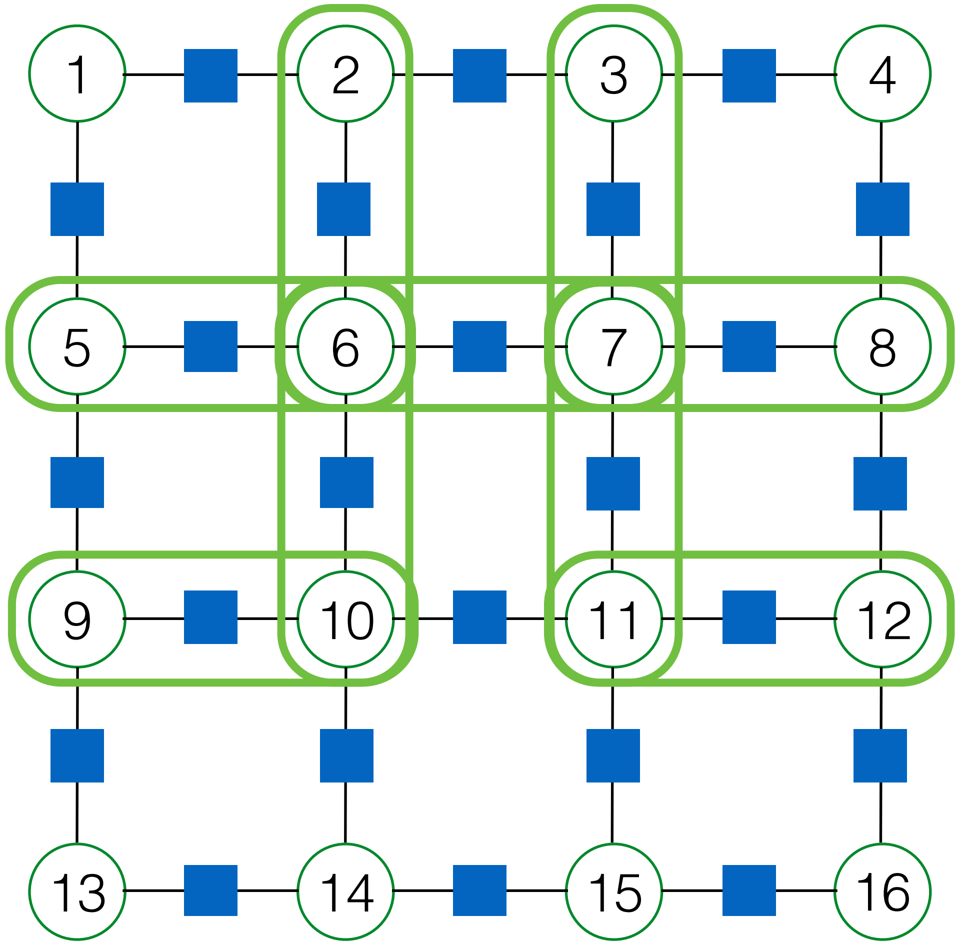
Region-Based Approximation

(The Choice of Regions)



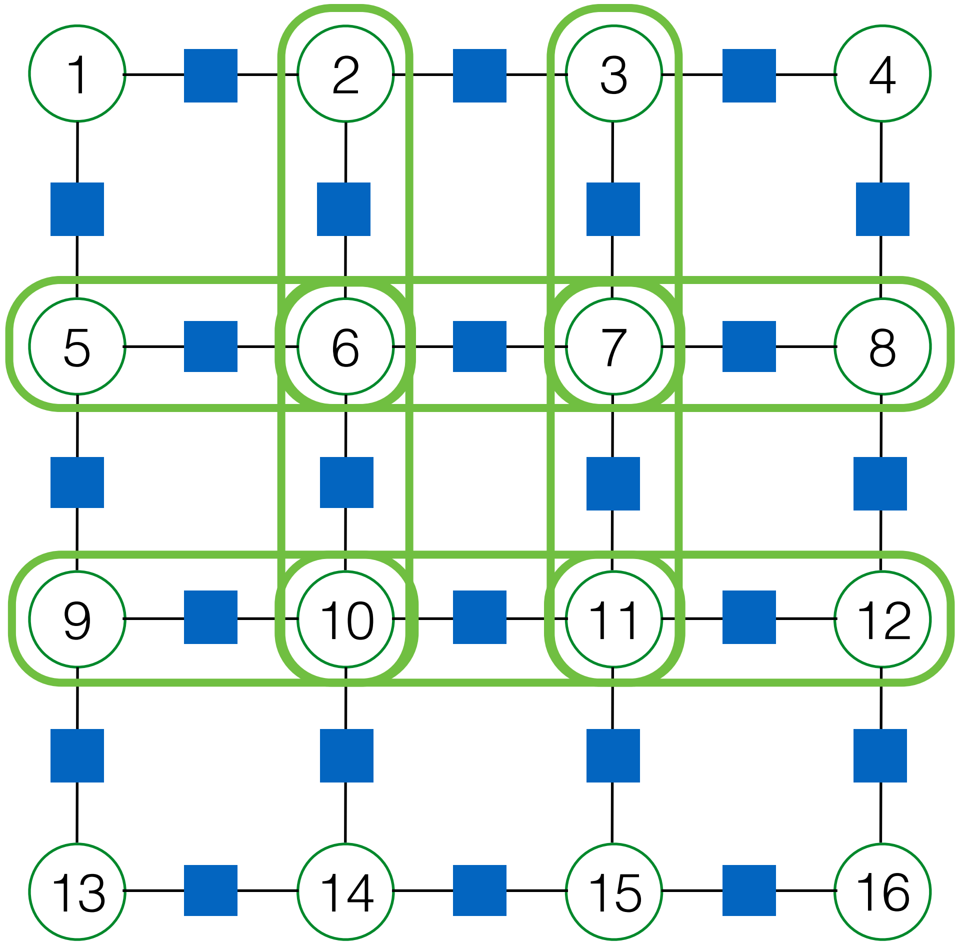
Region-Based Approximation

(The Choice of Regions)



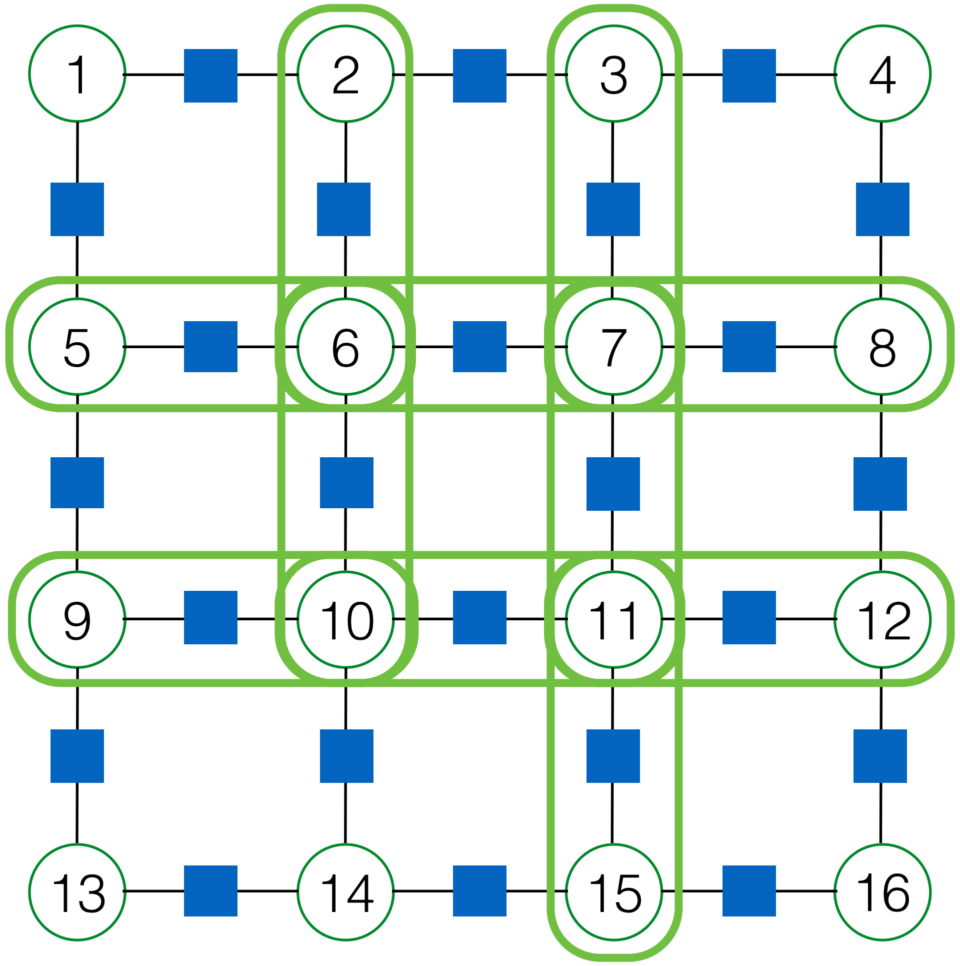
Region-Based Approximation

(The Choice of Regions)



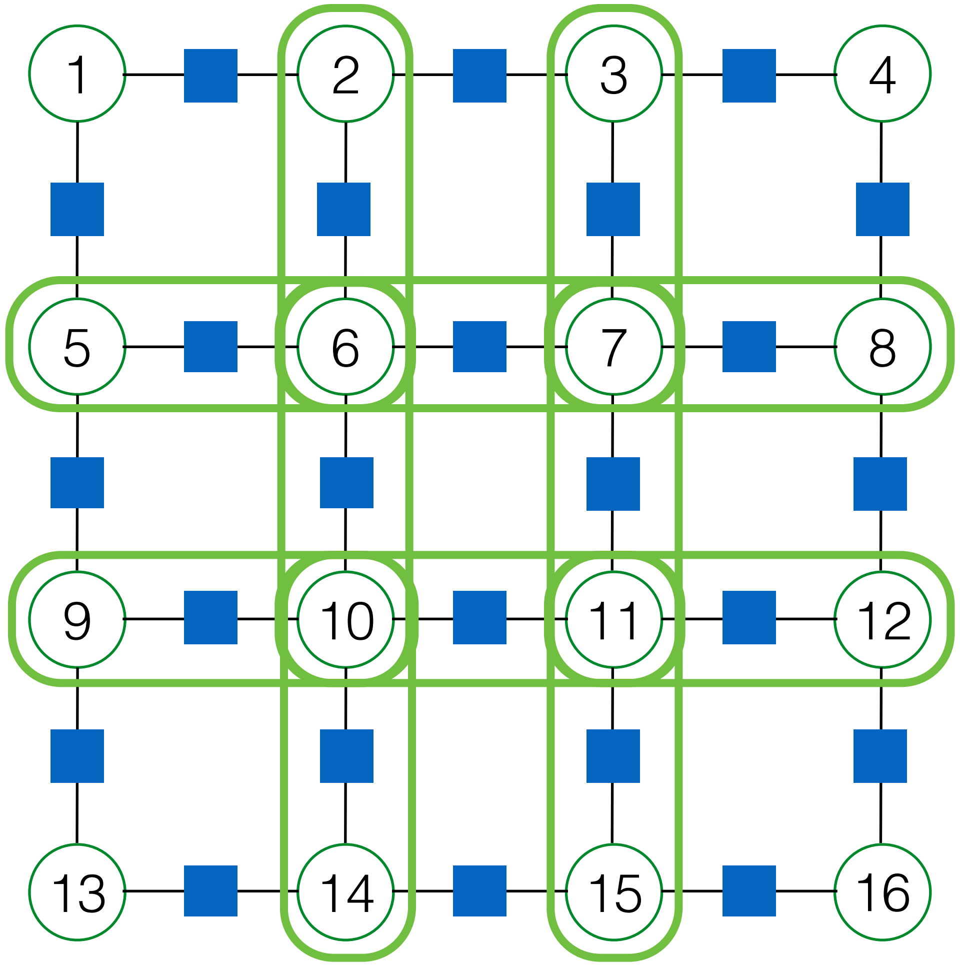
Region-Based Approximation

(The Choice of Regions)



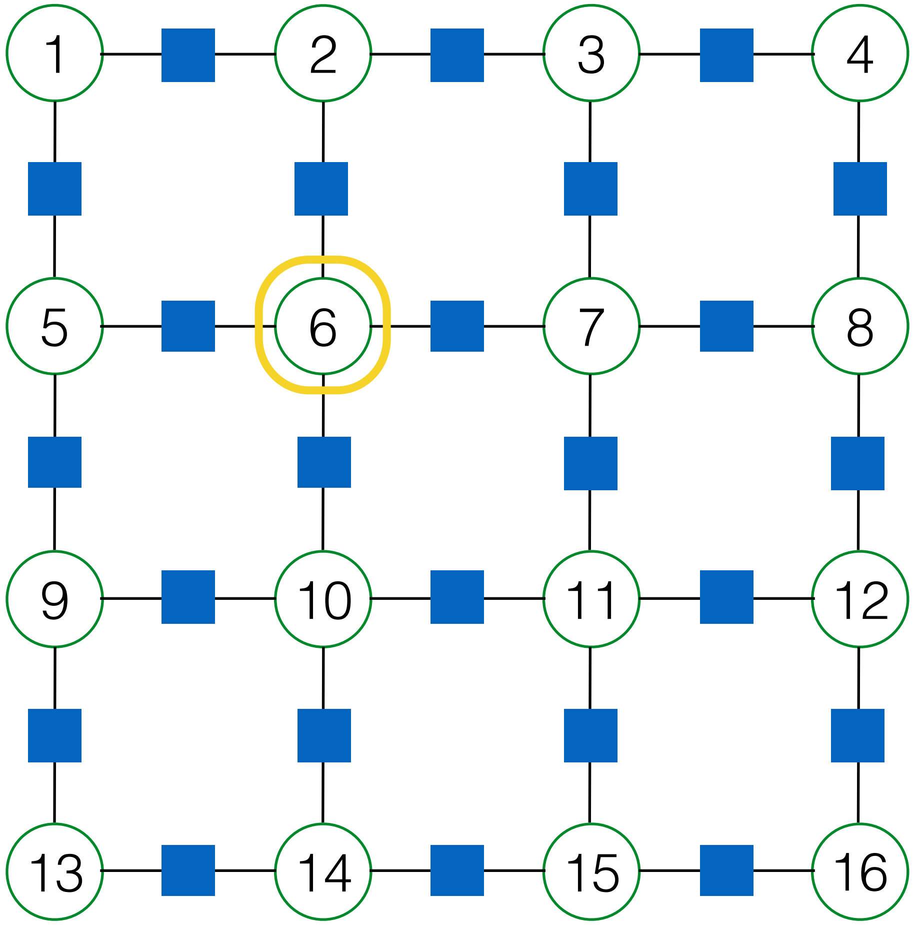
Region-Based Approximation

(The Choice of Regions)



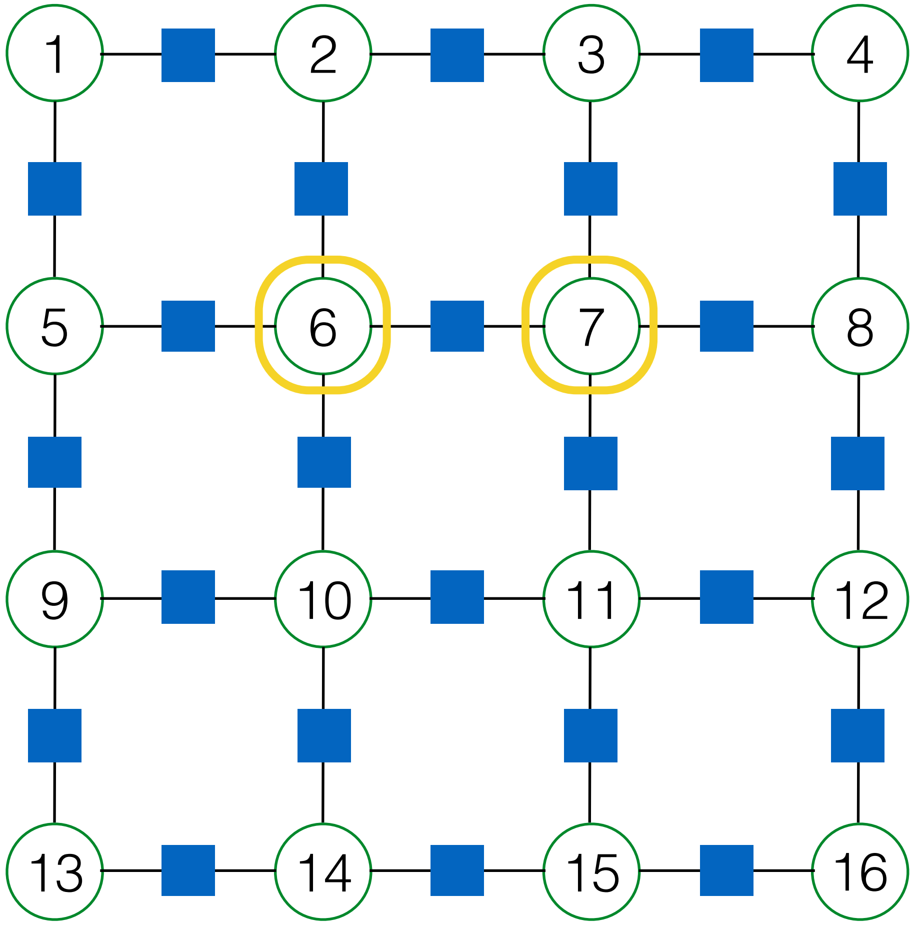
Region-Based Approximation

(The Choice of Regions)



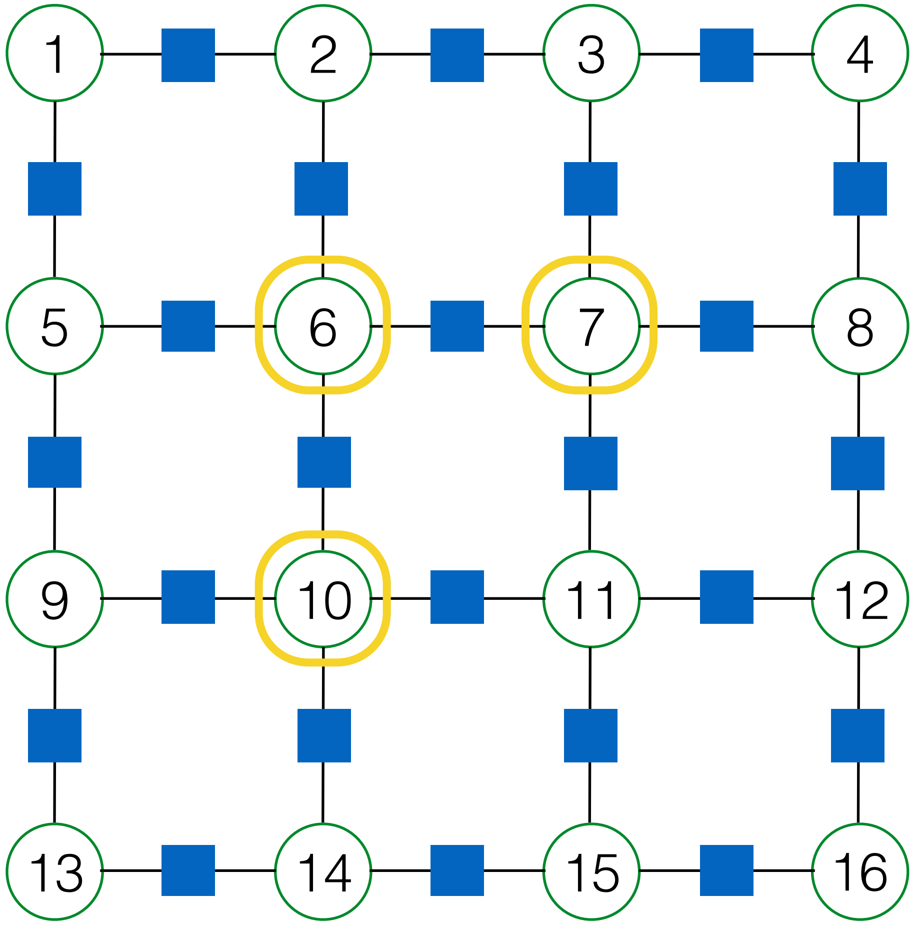
Region-Based Approximation

(The Choice of Regions)



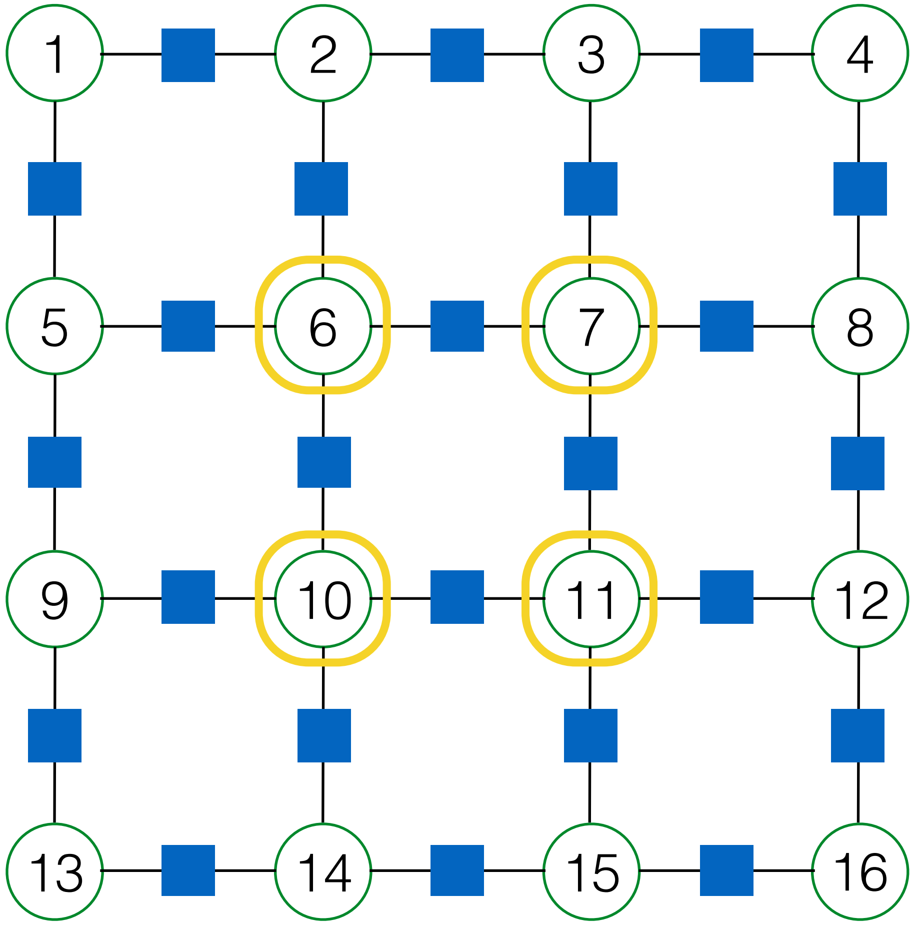
Region-Based Approximation

(The Choice of Regions)



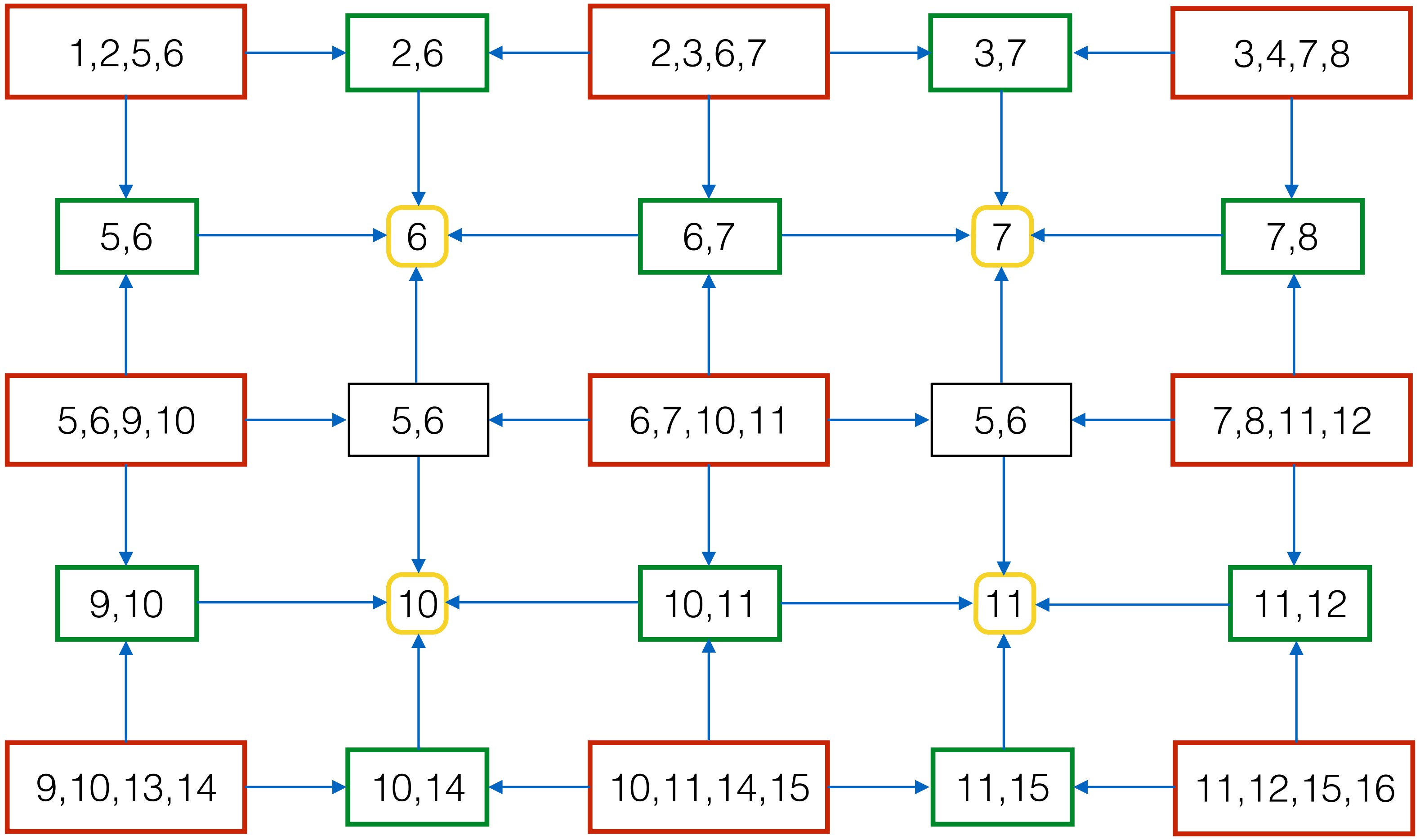
Region-Based Approximation

(The Choice of Regions)



Region-Based Approximation

(The Region Graph)



1,2,5,6 2,6 2,3,6,7 3,7 3,4,7,8

5,6 6 6,7 7 7,8

5,6,9,10 5,6 6,7,10,11 5,6 7,8,11,12

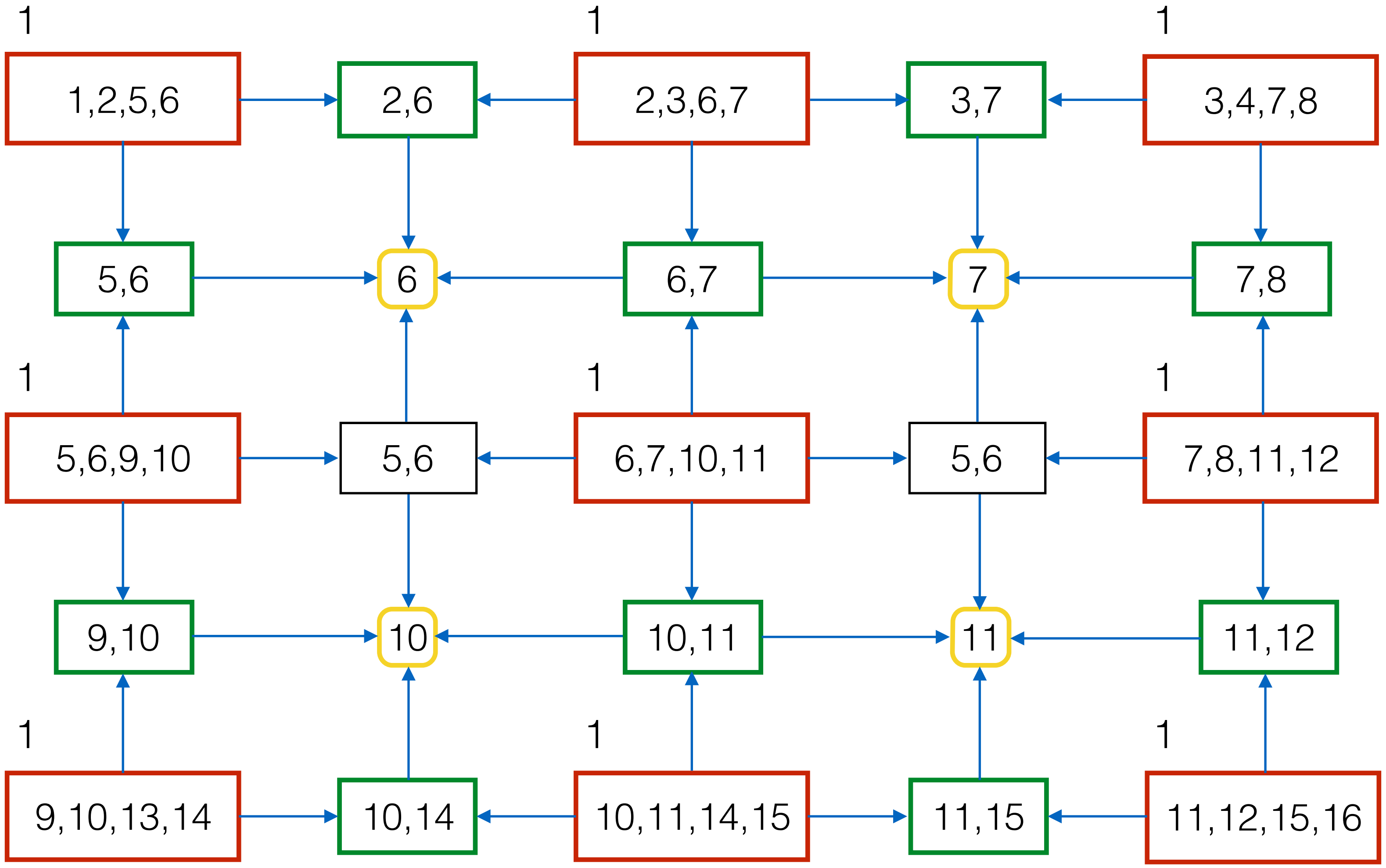
9,10 10 10,11 11 11,12

9,10,13,14 10,14 10,11,14,15 11,15 11,12,15,16

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R\in\mathcal{R}} c_R F_R(b_R)$$

Region-Based Approximation

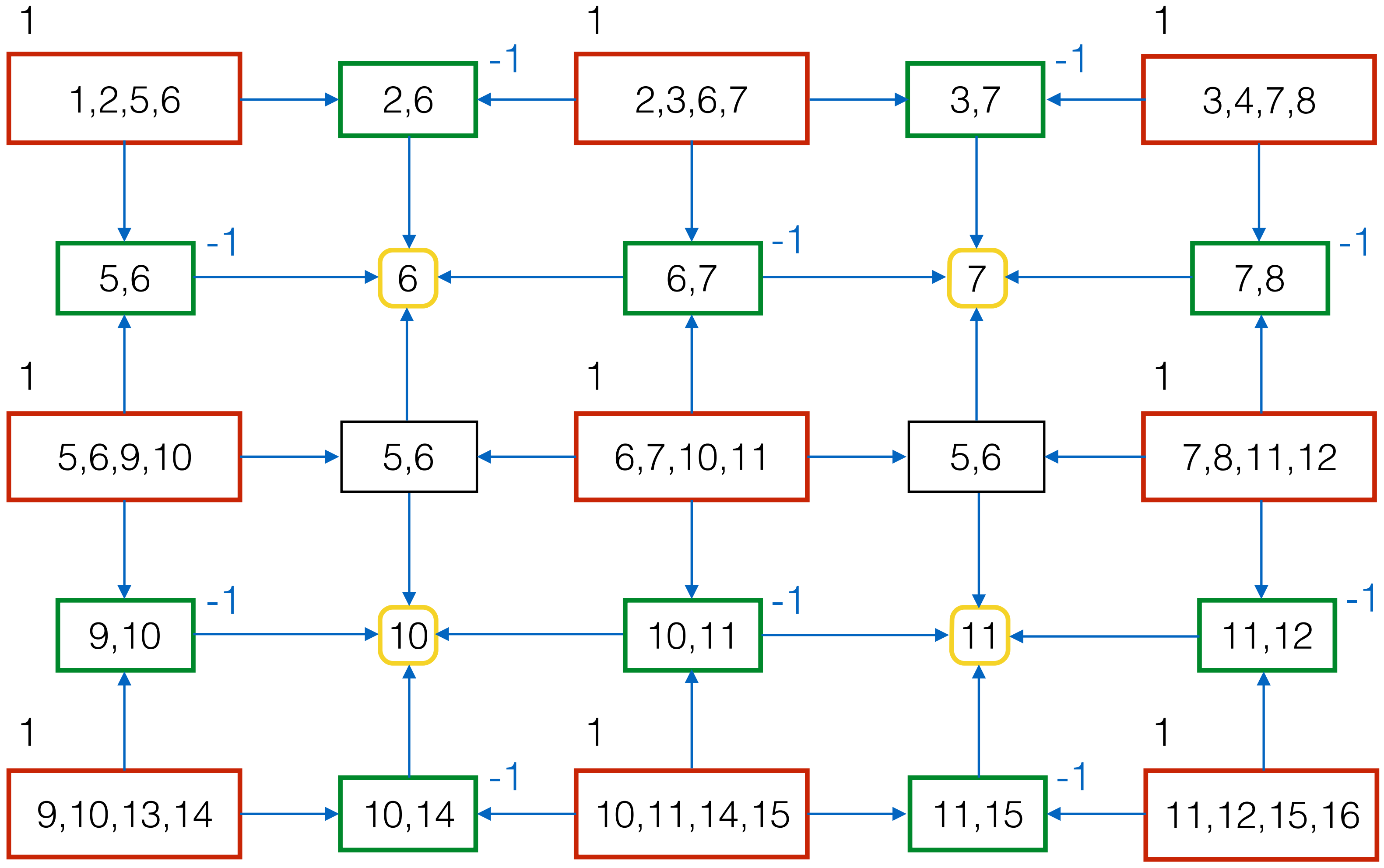
(The Region Graph)



$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R \in \mathcal{R}} c_R F_R(b_R)$$

Region-Based Approximation

(The Region Graph)



1 1 1

1,2,5,6 2,6 -1 2,3,6,7 3,7 -1 3,4,7,8

5,6 -1 6 6,7 -1 7 7,8 -1

1 1 1

5,6,9,10 5,6 6,7,10,11 5,6 7,8,11,12

9,10 -1 10 10,11 -1 11 11,12 -1

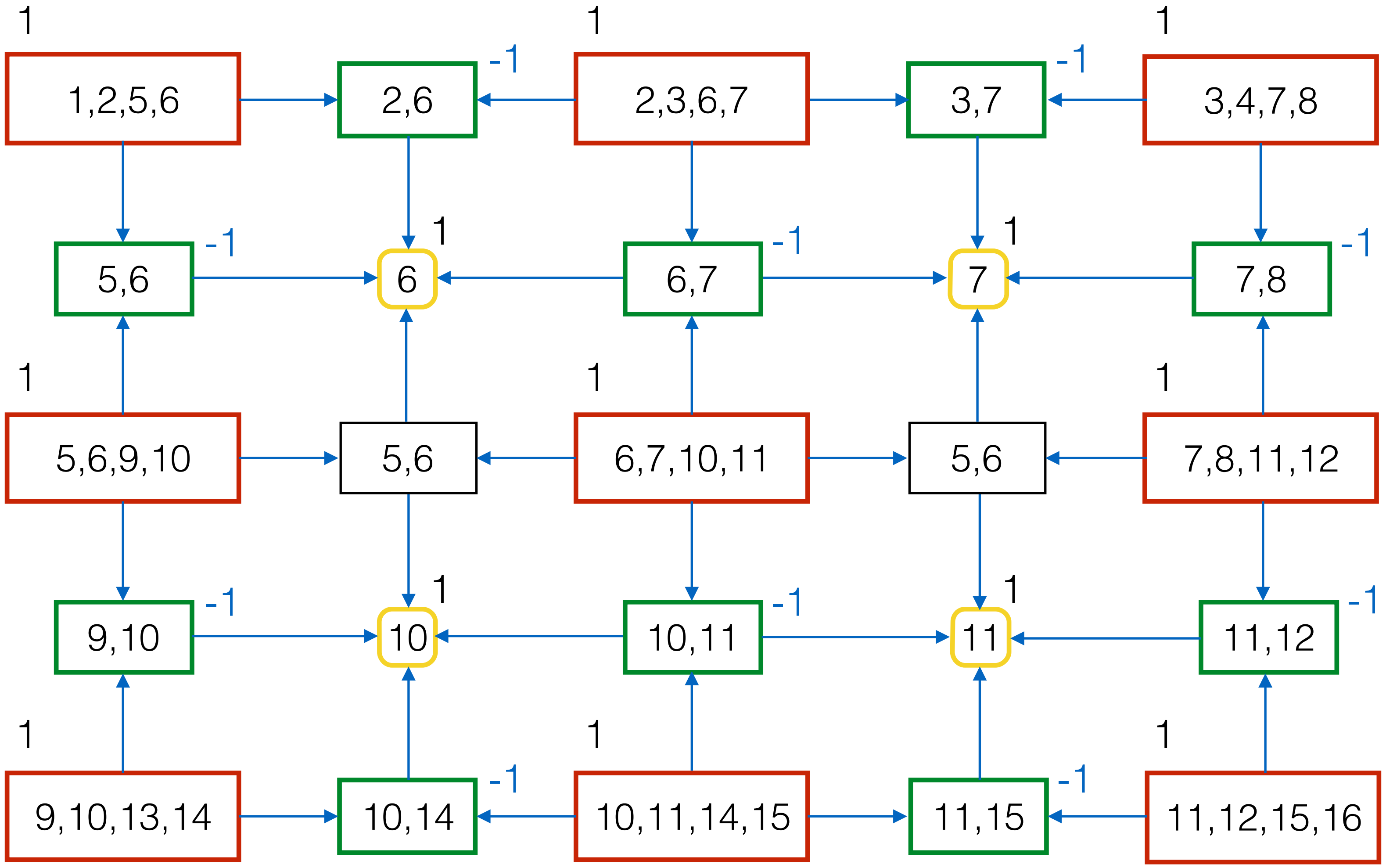
1 1 1

9,10,13,14 10,14 -1 10,11,14,15 11,15 -1 11,12,15,16

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R \in \mathcal{R}} c_R F_R(b_R)$$

Region-Based Approximation

(The Region Graph)



1 1,2,5,6
2,6 -1
1 2,3,6,7
3,7 -1
1 3,4,7,8

5,6 -1
6 1
6,7 -1
7 1
7,8 -1

1 5,6,9,10
5,6
1 6,7,10,11
5,6
1 7,8,11,12

9,10 -1
10 1
10,11 -1
11 1
11,12 -1

1 9,10,13,14
10,14 -1
1 10,11,14,15
11,15 -1
1 11,12,15,16

$$F_{\mathcal{R}}(\{b_R\}) = \sum_{R\in\mathcal{R}} c_R F_R(b_R)$$

Previous Work and Our Result

- Previous work:
  For any binary log-supermodular graphical model, for any fixed pound of BP, we have

$$Z \geq Z_{\mathrm{BP}}\big(\{b_i, b_a\}\big). \qquad F_{\mathrm{B}}(\{b_i, b_a\}) = -\log Z_{\mathrm{BP}}\big(\{b_i, b_a\}\big)$$

- Our result:
  For $R_{m\times n}$ based on 2D Ising model of size no large than 5 x 5 or 3 x n, for any fixed pound of GBP, we have

$$Z \geq Z_{\mathcal{R},\mathrm{GBP}}\big(\{b_R\}\big). \qquad F_{\mathcal{R}}(\{b_R\}) = -\log Z_{\mathcal{R},\mathrm{GBP}}\big(\{b_R\}\big)$$

- Conjecture:
  The above statement is true for any $R_{m\times n}$ based on 2D Ising model of any size

Proof Idea

- First, we show that

$$\frac{Z}{Z_{\mathcal{R},\mathrm{GBP}}(\{b_R\})} = \sum_{\boldsymbol{x}} \prod_{R\in\mathcal{R}} (b_R(\boldsymbol{x}_R))^{c_R}$$

- Using result of Ruozzi, we can show that the 2D Ising model can be transformed to a log-supermodular graphical model
    - This transformation preserves the partition function and also does not change the fixed-point-based approximation of partition function using GBP
- Next, we analyze the above ratio for binary pairwise graphical models with log-supermodular factor function

Thank You!

Some of the References

- J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," IEEE Trans. Inf. Theory, 2005.

- E. B. Sudderth, M. Wainwright, and A. S. Willsky, "Loop series and Bethe variational bounds in attractive graphical models," NIPS 2007.

- N. Ruozzi, "The Bethe partition function of log-supermodular graphical models," NIPS, 2012.

- Sabato, G. and Molkaraie, M., "Generalized Belief propagation for the noiseless capacity and information rates of run-length limited constraints," IEEE Trans. Comm., 2012.

- C. L. Chan, M. J. Siavoshani, S. Jaggi, N. Kashyap, and P. O. Vontobel, "Generalized Belief Propagation for Estimating the Partition Function of the 2D Ising Model," ISIT'15, Hong Kong, 2015.