Structured Region Graphs: Morphing EP into GBP

Max Welling

Tom Minka

Yee Whye Teh

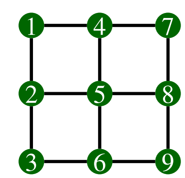
GBP and EP

- Approximate inference in large graphical models
- Generalized belief propagation [Yedidia, Freeman, Weiss, NIPS 2000]
  - Minimize Kikuchi free energy
- Expectation propagation [Minka, UAI 2001]
  - Minimize local KL-divergence
- Require choosing approximation structure
  - Kikuchi clusters, exponential family
- Need a constructive framework…

Structured Region Graphs

- A general representation for both GBP and EP approximations
- Reveals equivalence between GBP/EP
  - Can convert between equivalent GBP/EP algorithms
- Simple tests ensure good performance: non-singularity, $\sum_R c_R = 1$, maximality
- A framework for constructing good SRGs for any graphical model

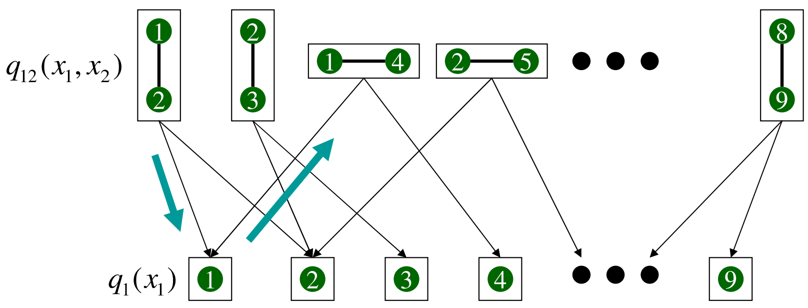
A simple graphical model



$$p(\mathbf{x}) = f_{12}(x_1, x_2) f_{23}(x_2, x_3) f_{14}(x_1, x_4) f_{25}(x_2, x_5) \cdots$$

Want single-variable marginals $p(x_1)$, $p(x_2)$, …

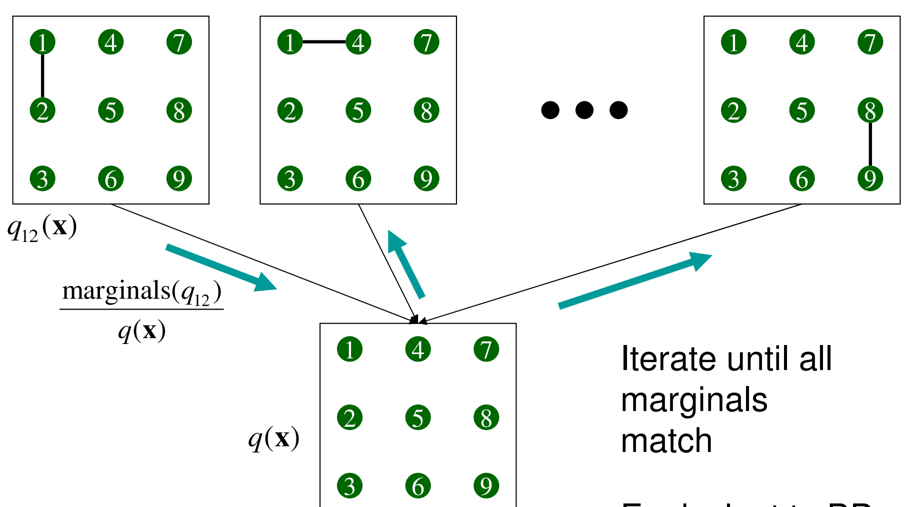
Belief propagation



$$\Delta msg_{12\to1}(x_1) = \frac{\sum_{x_2} q_{12}(x_1, x_2)}{q_1(x_1)} = \Delta msg_{1\to14}(x_1)$$

Iterate until all marginals match

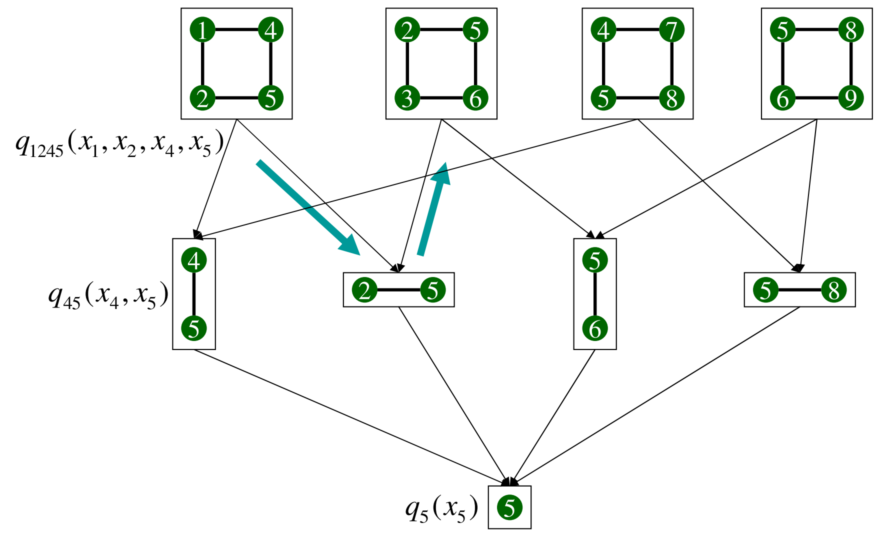
Fully-factorized EP



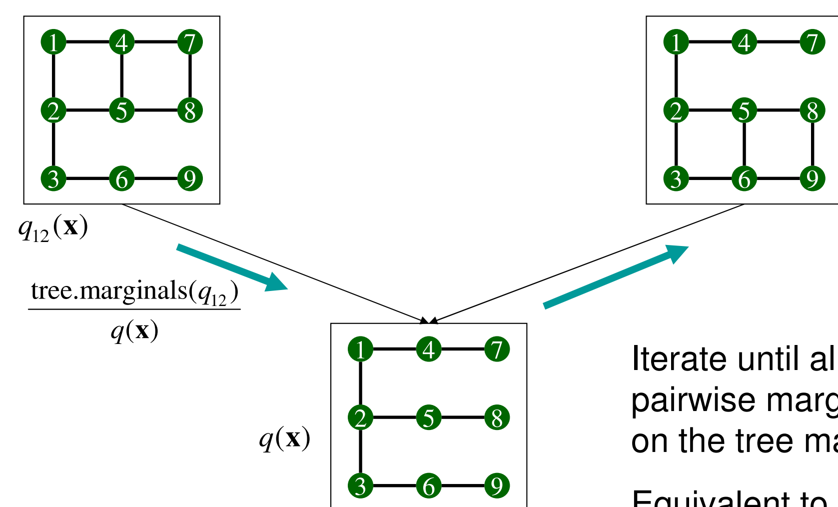
Iterate until all marginals match

Equivalent to BP

Generalized belief propagation



Tree-structured EP



Iterate until all pairwise marginals on the tree match

Equivalent to GBP on squares

Common theme

- GBP and EP approximate p(x) in a distributed fashion
- Factors are allocated to local regions
- Each region has a distribution of a specific form, tied together by constraints
- Regions pass messages until they meet the constraints

Approximation choices

1. Number of regions
2. Allocation of factors to regions
3. Number of parameters per region
4. Which regions to constrain
5. What type of constraints

- How can we reason about these choices?

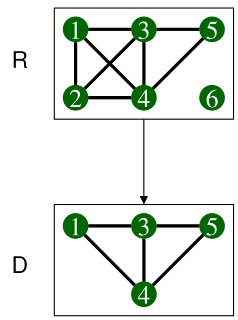
Outline

- Structured region graphs
- Equivalence operators
- Design criteria
- Design examples

Structured Region Graph

- A general representation for GBP and EP approximations
- A DAG of regions, each with a graph structure, and a set of factors
- Graph structure defines the form of $q_R(x_R)$
- Links define constraints – parent and child have the same clique-marginals
- Extends region graph formalism of [Yedidia,Freeman,Weiss, 2002]

Structured Region Graph



Cliques (1,3,4)(3,4,5)

q_R must match q_D on (1,3,4) and (3,4,5):

$$\sum_{\mathbf{x} \setminus (x_1, x_3, x_4)} q_R(\mathbf{x}_R) = q_D(x_1, x_3, x_4)$$

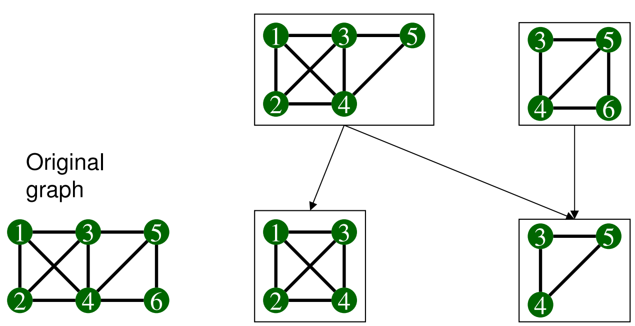
$$\sum_{\mathbf{x} \setminus (x_3, x_4, x_5)} q_R(\mathbf{x}_R) = q_D(x_3, x_4, x_5)$$

Parent must be super-graph of child

GBP region graphs

- All inner regions are complete
  [Yedida,Freeman,Weiss, 2002]
- Thus $q_R(x_R)$ is not factorized

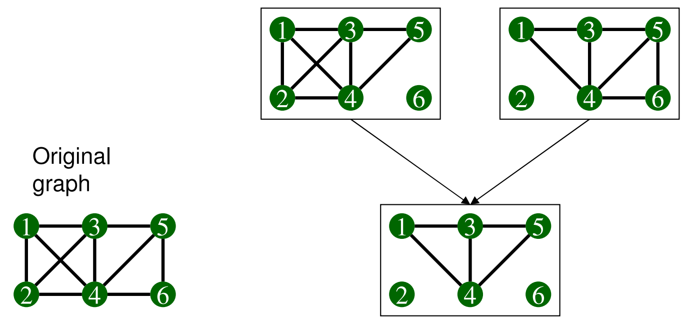
Original graph



EP region graphs

- Only one inner region
- Every region contains all variables

Original graph



Free energy

- Each region has counting number

$$c_R = 1 - \sum_{A \in \mathrm{an}(R)} c_A$$

- Free energy:

$$F(q \,\|\, p) = \sum_R c_R \sum_{x_R} q_R(x_R) \log \frac{q_R(x_R)}{f_R(x_R)}$$

subject to the parent-child marginal constraints

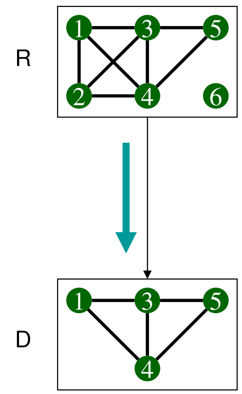
Applies to both GBP and EP (special cases)

Generalized EP messages

- Parent-child algorithm (for discrete variables):

$$\Delta msg_{R \to D}(\mathbf{x}_D) = \frac{\text{clique.marginals}(q_R)}{q_D(\mathbf{x}_D)}$$

- D relays this to other parents
- Iterate until all constraints satisfied
- Fixed point of msg passing = critical point of free energy



Outline

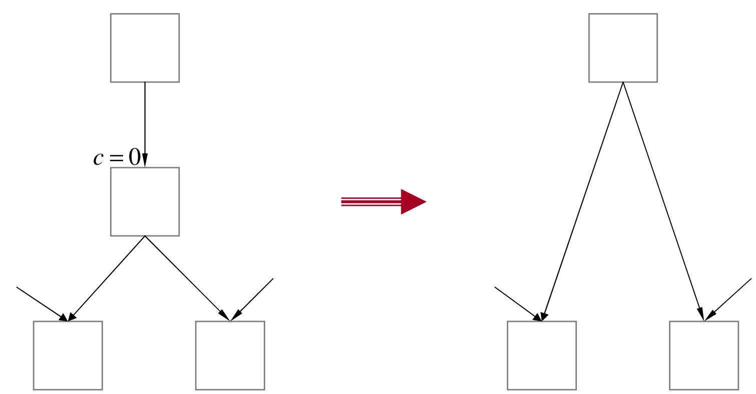
- Structured region graphs
- **Equivalence operators**
- Design criteria
- Design examples

Equivalence operators

- Graphical operators that preserve the critical points of the free energy:
  1. **Region-Drop**
  2. **Region-Merge**
  3. **Region-Split**
  4. **Link-Death**
  5. **Clique-Grow/Shrink**
  6. **Factor-Move**

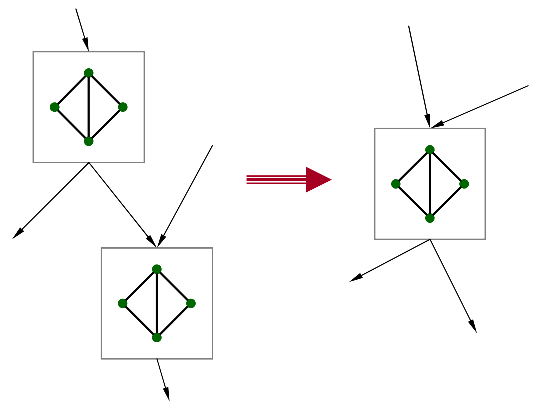
Region Drop

- A region with one parent can be dropped (replaced by direct links)



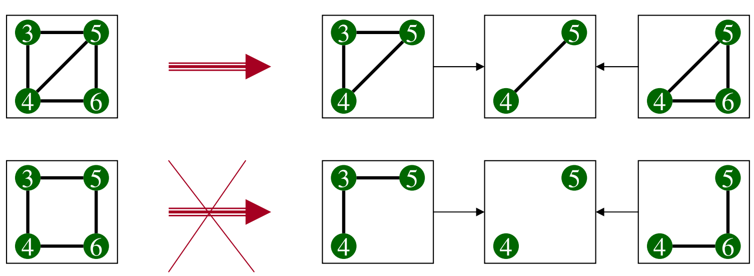
Region Merge

- Linked regions with the same structure can be merged



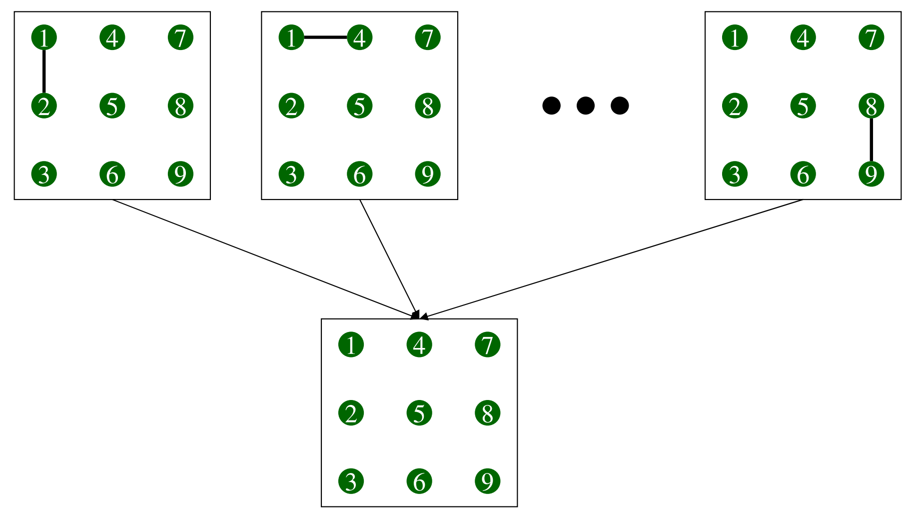
Region split

- Any region can be split into two regions plus a separator
- Separator must be complete
- Pieces must be super-graphs of children

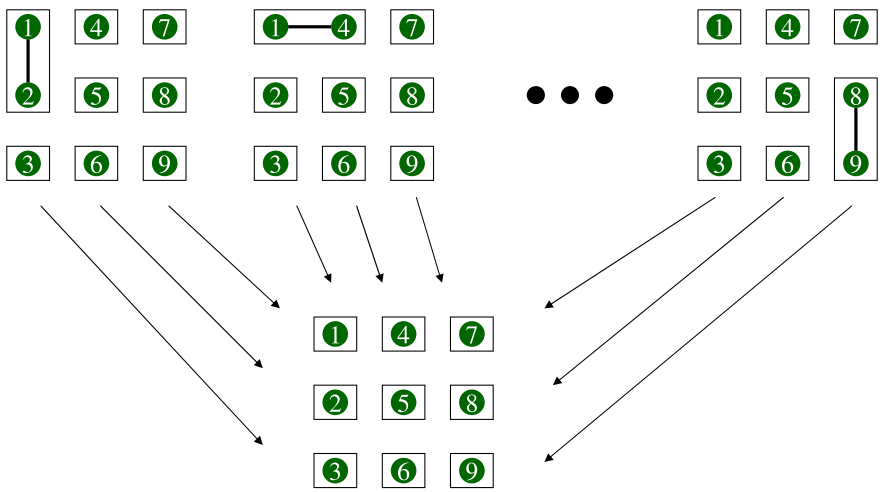


Equivalence of BP and fully-factorized EP

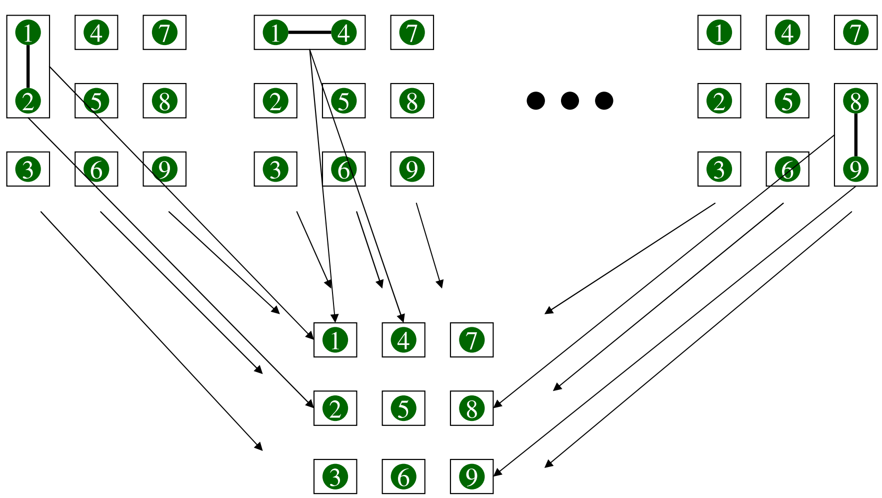
Fully-factorized EP



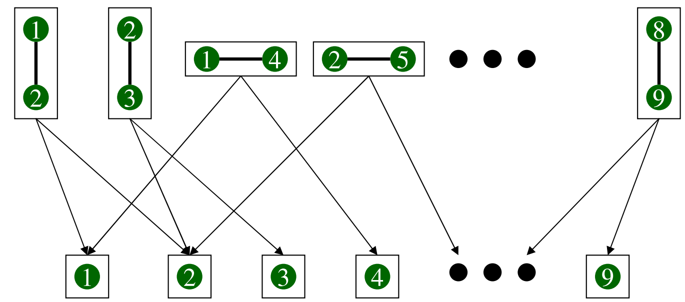
SPLIT



MERGE



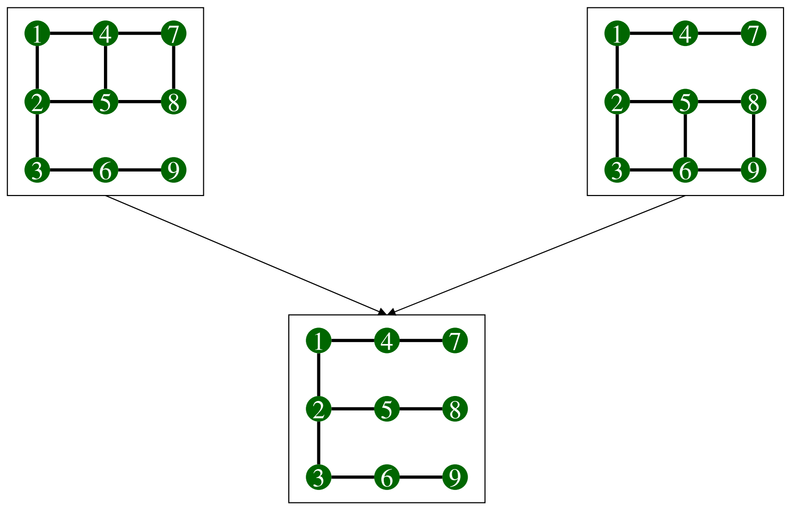
Belief propagation graph



BP and fully-factorized EP have the same fixed points

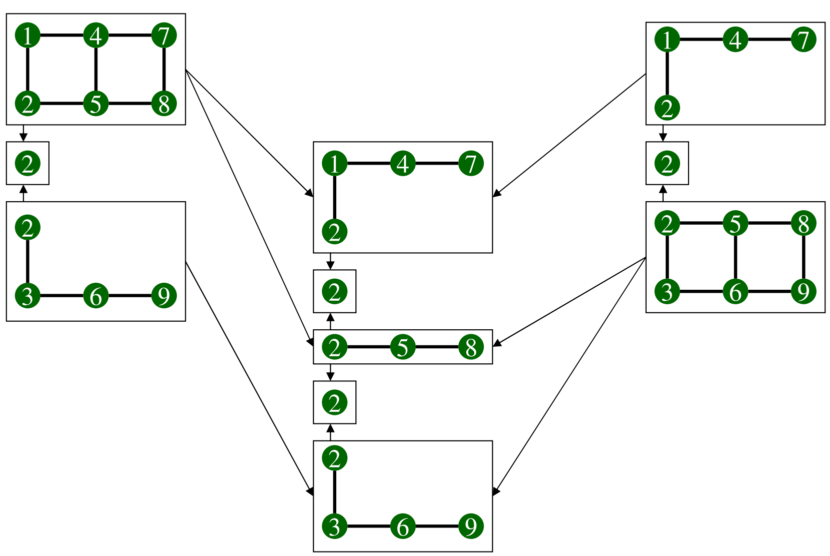
Equivalence of GBP-squares and tree-structured EP

Tree-structured EP



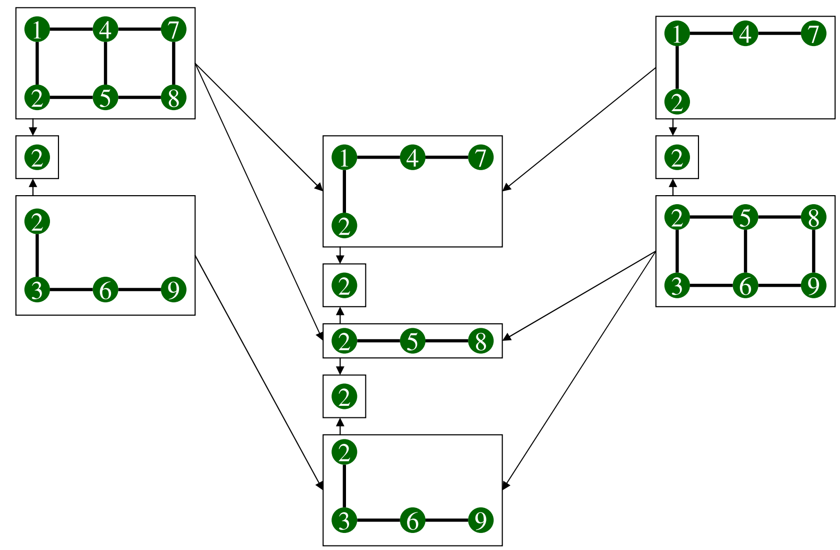
29

SPLIT



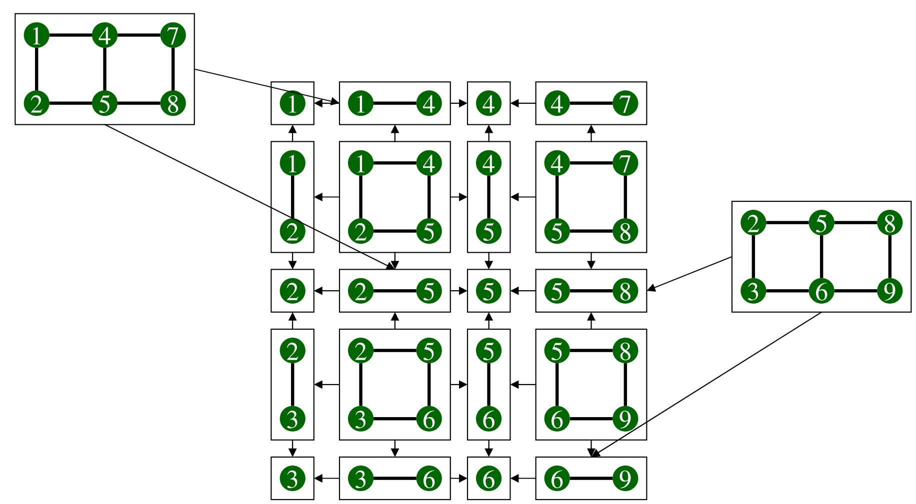
30

MERGE



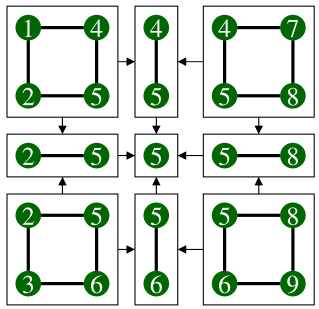
31

SPLIT

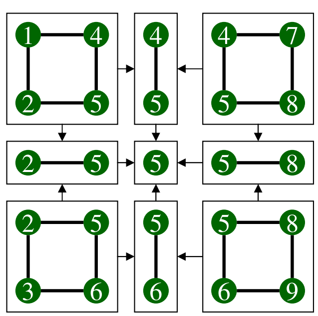


32

DROP



GBP-squares region graph



- The chosen TreeEP region graph has the same fixed points as GBP-squares
- Extends to any grid

When does EP reduce to GBP?

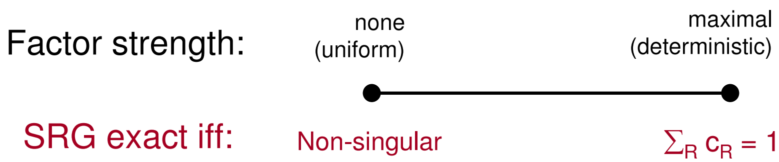
- When all variables are discrete, and inner region is triangulated (i.e. approximation family is decomposable)
- E.g. TreeEP always reduces to GBP
- Proof: split all inner regions, starting at the bottom, until only complete regions are left
- But EP is often faster
  - (10x faster in [Minka & Qi, NIPS 2003])

Outline

- Structured region graphs
- Equivalence operators
- **Design criteria**
- Design examples

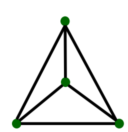
Good region graphs

- Consider 2 extreme cases:
  - maximally correlated variables (strong factors)
  - uniform variables (weak factors)
- Want approx to be exact in (at least) these cases [Yedidia,Freeman,Weiss, 2004]



Non-singularity

- Def: All fixed points are uniform when the factors are uniform
- Not true for all region graphs
- Equivalent def: No ‘redundant’ regions
  - create spurious fixed points
  - analogous to singular matrix
- E.g. all triples in K_4 = singular

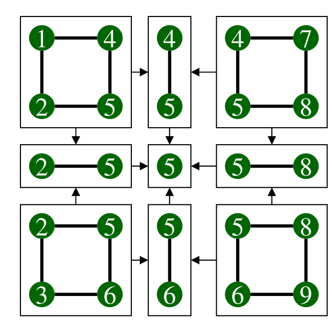


Simple test for non-singularity

- Non-singularity is preserved by equivalence operators
- Theorem: SRG is non-singular iff reduces to single-variable regions when all factors are removed

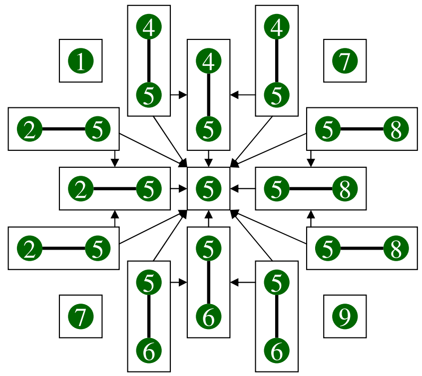
Example: Squares graph

1. Remove factors



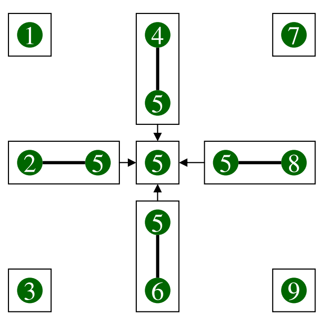
Example: Squares graph

1. Remove factors
2. Split



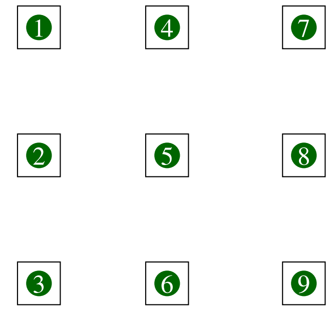
Example: Squares graph

1. Remove factors
2. Split
3. Merge
4. Clique-shrink



Example: Squares graph

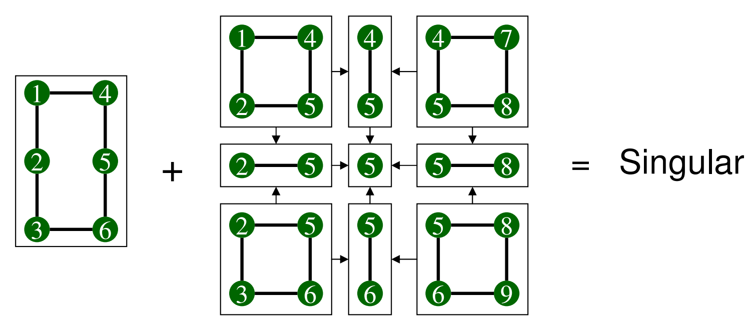
1. Remove factors
2. Split
3. Merge
4. Clique-shrink
5. Split & merge



The squares graph is non-singular

Example: An extra loop

- Adding *any* extra loop (and overlap edges) to the squares graph makes it singular
- Squares graph is *maximal wrt loops*



General results

- Every acyclic SRG (no cycles of regions) is non-singular and has $\sum_R c_R = 1$
  - EP-graphs are acyclic
- If all regions contain at most one loop, then non-singular & $\sum_R c_R = 1$ implies *maximal wrt loops*
  - E.g. squares graph

Outline

- Structured region graphs
- Equivalence operators
- Design criteria
- Design examples

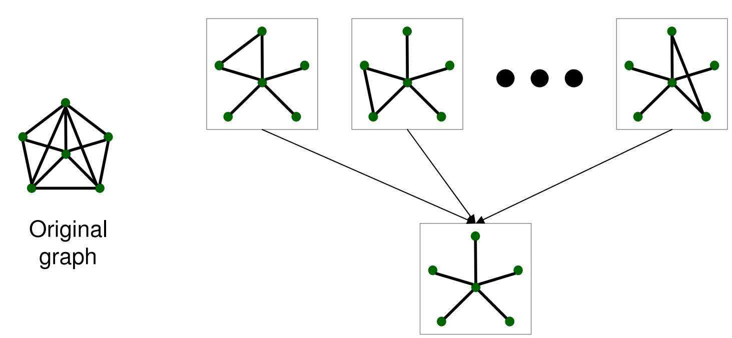
Region graph design

- Want non-singular, $\sum_R c_R = 1$, maximal

1. Start with EP-graph and reduce
2. Start with BP-graph and add regions (region pursuit)

Star graph

- Non-singular, $\sum_R c_R = 1$, maximal
- Closed under intersection
- Very effective on dense graphs

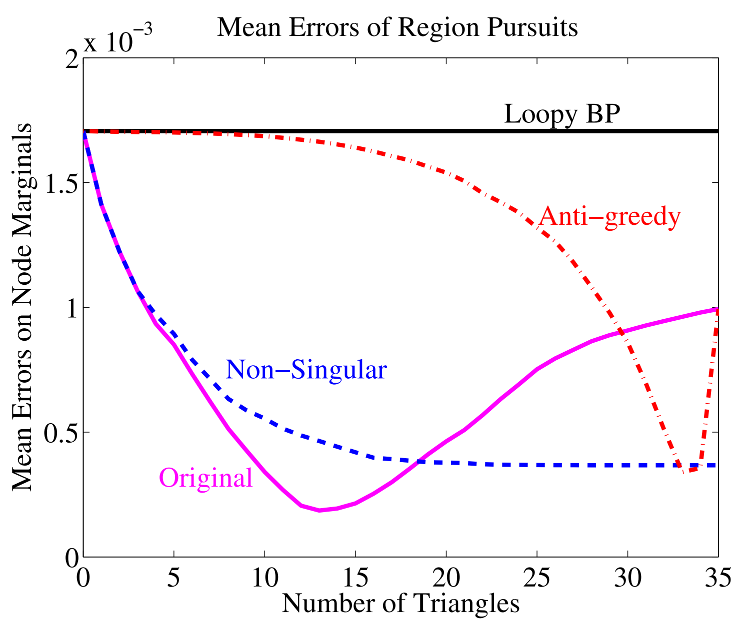


Region pursuit

- Start with edge regions only
- Greedily add the most "significant" cluster
  - changes free energy the most

  [Welling, UAI 2004]
- Performs poorly when too many clusters are added
- New twist: Skip clusters which would make the graph singular (tested automatically)

7-node complete graph



Summary

- A general formalism for GBP and EP approximations
- Equivalence operators between SRGs
  - equivalences between EP and GBP
- Simple tests ensure good performance: non-singularity, $\sum_R c_R = 1$, maximality

Future work

- More design principles
  - strength of actual factors
  - closed under intersection
- General test for maximality
- Generalized EP on continuous variables [Heskes & Zoeter, AISTATS 2003]