Revisiting the Limits of MAP Inference by MWSS on Perfect Graphs

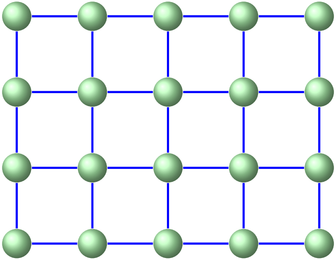
Adrian Weller
University of Cambridge

CP 2015
Cork, Ireland

Slides and full paper at
http://mlg.eng.cam.ac.uk/adrian/

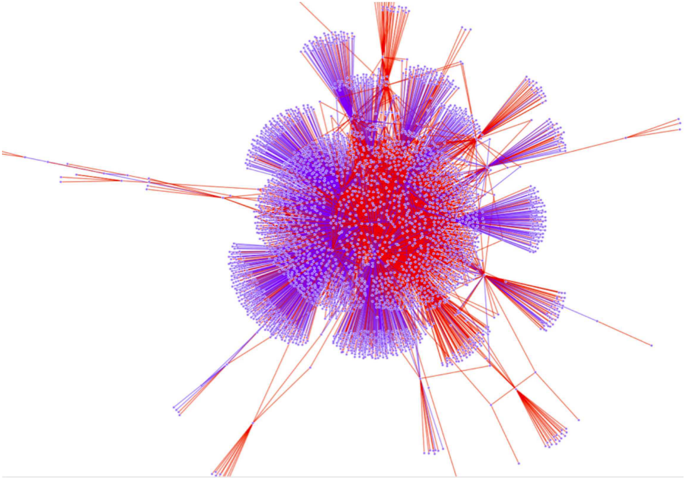
Motivation: *undirected graphical models (MRFs)*

- Powerful way to represent relationships across variables
- Many applications including: computer vision, social network analysis, deep belief networks, protein folding...
- In this talk, mostly focus on binary pairwise (Boolean binary or *Ising*) models



Example: Grid for computer vision (attractive)

Motivation: *undirected graphical models*



Example: epinions social network (attractive and repulsive edges)

Figure courtesy of N. Ruozzi

Motivation: *undirected graphical models*

A fundamental problem is *maximum a posteriori (MAP) inference*

- Find a global mode configuration with highest probability

$$x^* \in \operatorname*{arg\,max}_{x=(x_1,\ldots,x_n)} p(x_1, x_2, \ldots, x_n)$$

- In a graphical model,

$$p(x_1, x_2, \ldots, x_n) \propto \exp\left(\sum_{c \in C} \psi_c(x_c)\right)$$

  where each c is a subset of variables, x_c is a configuration of those variables, and $\psi_c(x_c) \in \mathbb{Q}$ is a *potential function*.

- Each potential function assigns a score to each configuration of variables in its scope, higher score for higher compatibility. May be considered a ‘negative cost’ function.

Motivation: *undirected graphical models*

A fundamental problem is *maximum a posteriori (MAP) inference*

- Find a global mode configuration with highest probability

$$x^* \in \underset{x=(x_1,\ldots,x_n)}{\arg\max} \sum_{c \in C} \psi_c(x_c), \quad \text{all } \psi_c(x_c) \in \mathbb{Q}$$

- Equivalent to finding a minimum solution of a *valued constraint satisfaction problem* (VCSP) without hard constraints: $x^* \in \arg\min_{x=(x_1,\ldots,x_n)} \sum_{c \in C} -\psi_c(x_c)$
- We are interested in *when is this efficient?* i.e. solvable in time polynomial in the number of variables

Overview of the method (for models of any arity)

We explore the limits of an exciting recent method (Jebara, 2009):

- Reduce the problem to finding a *maximum weight stable set* (MWSS) in a derived weighted graph called a *nand Markov random field* (NMRF)
- Examine how to *prune* the NMRF (removes nodes, simplifies the problem)
- Different *reparameterizations* lead to pruning different nodes
- This allows us to solve the original MAP inference problem efficiently if some pruned NMRF is a *perfect graph*

Background: NMRFs and reparameterizations

- In the constraint community, an NMRF is equivalent to the complement of the microstructure of the dual representation (Jégou, 1993; Larrosa and Dechter, 2000; Cooper and Živný, 2011; El Mouelhi et al., 2013)
- Reparameterizations here are equivalent to considering soft arc consistency

A *reparameterization* is a transformation of potential functions (shifts score between potentials)

$$\{\psi_c\} \to \{\psi'_c\} \quad \text{s.t.} \quad \forall x, \sum_{c \in C} \psi_c(x_c) = \sum_{c \in C} \psi'_c(x_c)$$

This clearly does not modify our MAP problem

$$x^* \in \underset{x=(x_1,\dots,x_n)}{\arg\max} \sum_{c \in C} \psi_c(x_c) = \underset{x=(x_1,\dots,x_n)}{\arg\max} \sum_{c \in C} {\psi'}_c(x_c)$$

but can be helpful to simplify the problem after *pruning*.

Summary of results

Only a few cases were known always to admit efficient MAP inference, including:

- Acyclic models (via dynamic programming) STRUCTURE
- Attractive models, i.e. all edges attractive/submodular (via graph cuts or LP relaxation) LANGUAGE $\{\psi_c\}$
  - generalizes to *balanced* models (no *frustrated cycles*)

These were previously shown to be solvable via a perfect pruned NMRF. Here we establish the following limits, which characterize precisely the power of the approach using a hybrid condition:

Theorem (main result)

A binary pairwise model maps efficiently to a perfect pruned NMRF for any valid potentials iff each *block* of the model is balanced or *almost balanced*.

Frustrated, balanced, almost balanced

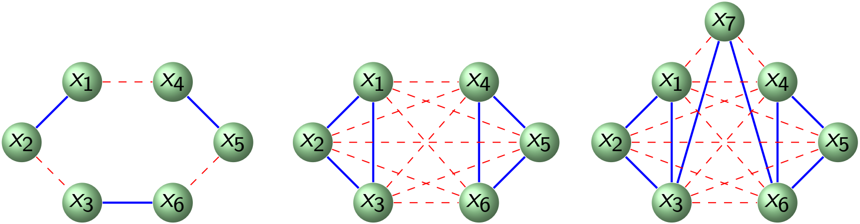
Each edge of a binary pairwise model may be characterized as:
- attractive (pulls variables toward the same value, equivalent to ψ_{ij} being supermodular or the cost function being submodular); or
- repulsive (pushes variables apart to different values).

- A *frustrated cycle* contains an odd number of repulsive edges. These are challenging for many methods of inference.
- A *balanced* model contains no frustrated cycle $\Leftrightarrow$ its variables form two partitions with all intra-edges attractive and all inter-edges repulsive.
- An *almost balanced* model contains a variable s.t. if it is removed, the remaining model is balanced.

Note all balanced models (with ≥ 1 variable) are almost balanced.

Examples: frustrated cycle, balanced, almost balanced

Signed graph topologies of binary pairwise models, solid blue edges are attractive, dashed red edges are repulsive:



frustrated cycle
(odd # repulsive edges)

balanced
(no frustrated cycle
so forms two partitions)

almost balanced
(added x_7)

a balanced model may be rendered attractive by 'flipping' all variables in one or other partition

Block decomposition

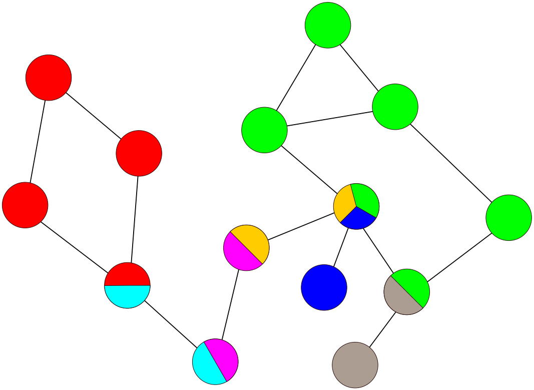


Figure from Wikipedia

Each color indicates a different block.

A graph may be repeatedly broken apart at *cut vertices* until what remains are the *blocks* (maximal 2-connected subgraphs).

Recap of result

Theorem (main result)

A binary pairwise model maps efficiently to a perfect pruned NMRF for any valid potentials iff each *block* of the model is *almost balanced*.

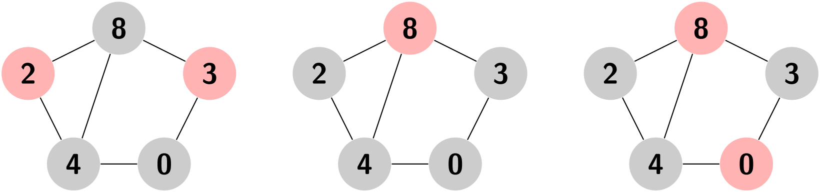
Note a model may have $\Omega(n)$ many blocks.

Next we discuss how to construct an NMRF and why the reduction works.

- We need some concepts from graph theory:
  - ▷ Stable sets, max weight stable sets (MWSS)
  - ▷ Perfect graphs

Stable sets, MWSS in weighted graphs

A set of (weighted) nodes is *stable* if there are no edges between any of them



Stable set | Max Weight Stable Set (MWSS) | Maximal MWSS (MMWSS)

- Finding a MWSS is NP-hard in general, but is known to be efficient for *perfect* graphs.

Perfect graphs

Perfect graphs were defined in 1960 by Claude Berge

- G is perfect iff $\chi(H) = \omega(H) \quad \forall$ induced subgraphs $H \leq G$
- Includes many important families of graphs such as bipartite and chordal graphs
- Several problems that are NP-hard in general, are solvable in polynomial time for perfect graphs: MWSS, graph coloring...
- We can use many known results, including:

  ▷ Strong Perfect Graph Theorem (Chudnovsky et al., 2006): G is perfect iff it contains no odd hole or antihole
  ▷ Pasting any two perfect graphs on a common clique yields another perfect graph

Reduction to MWSS on an NMRF

Recall our theme: Given a model, we construct a weighted graph NMRF. Claim: If we can solve MWSS on the NMRF, we recover a MAP solution to the original model.
If the NMRF is perfect, MWSS runs in polynomial time.

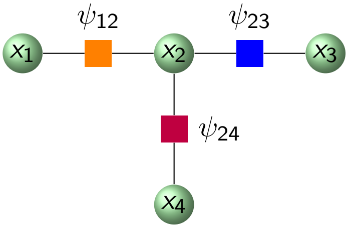
Idea: A MAP configuration has
$\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all the x_c are consistent, consistency will be enforced by requiring a stable set.

We construct a *nand Markov random field* (NMRF, Jebara, 2009; equivalent to the complement of the microstructure of the dual) N:

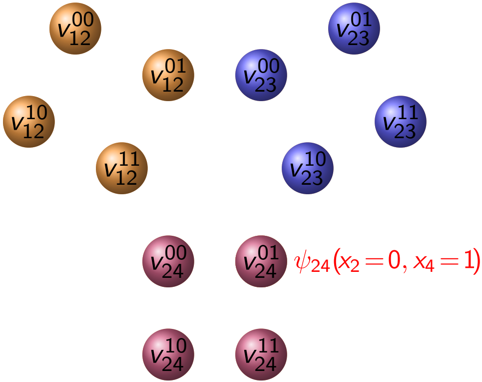
- For each potential ψ_c, instantiate a node in N for every possible configuration x_c of the variables in its scope c
- Give each node a weight $\psi_c(x_c)$ *then adjust*
- Add edges between any nodes which have inconsistent settings

Example: constructing an NMRF

Idea: A MAP configuration has $\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all x_c are consistent, consistency will be enforced by requiring a stable set.



Original model (factor graph)



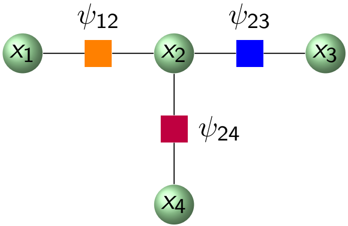
Derived NMRF

superscripts denote configuration x_c

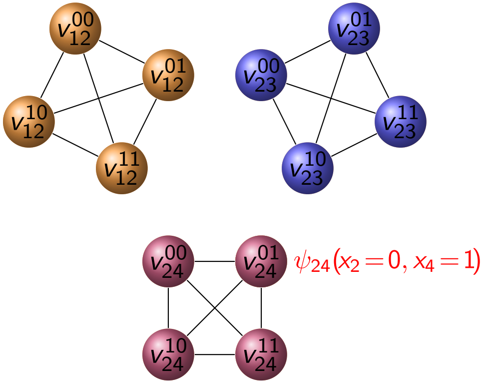
subscripts denote variable set c

Example: constructing an NMRF

Idea: A MAP configuration has $\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all x_c are consistent, consistency will be enforced by requiring a stable set.



Original model (factor graph)



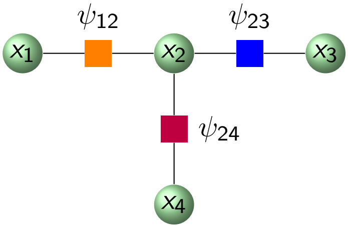
Derived NMRF

superscripts denote configuration x_c

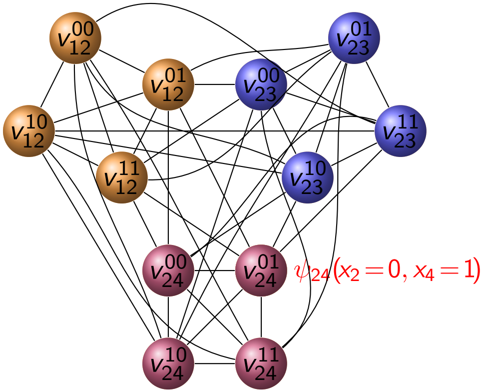
subscripts denote variable set c

Example: constructing an NMRF

Idea: A MAP configuration has $\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all x_c are consistent, consistency will be enforced by requiring a stable set.



Original model (factor graph)



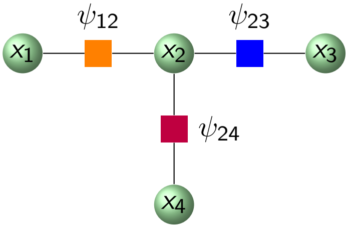
Derived NMRF

superscripts denote configuration x_c

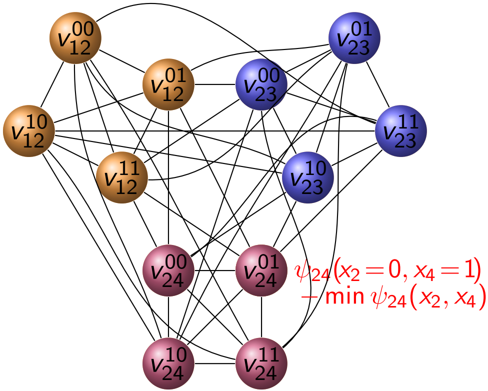
subscripts denote variable set c

Example: constructing an NMRF

Idea: A MAP configuration has $\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all x_c are consistent, consistency will be enforced by requiring a stable set.



Original model (factor graph)



Derived NMRF

superscripts denote configuration x_c

subscripts denote variable set c

Earlier results

Idea: A MAP configuration has
$\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all the x_c are consistent, consistency will be enforced by requiring a stable set.

- A MMWSS of the NMRF returns a MAP configuration of the original model.
- To find a MMWSS of the NMRF: zero-weight nodes may be *pruned* (removed), a MWSS found, then zero-weight nodes added back greedily.
- MAP inference is efficient if $\exists$ an efficiently identifiable efficient reparameterization s.t. the model maps to a perfect pruned NMRF.
- Decomposition: If each *block* of a model yields a perfect NMRF, then so too will the whole model (Weller and Jebara, 2013).

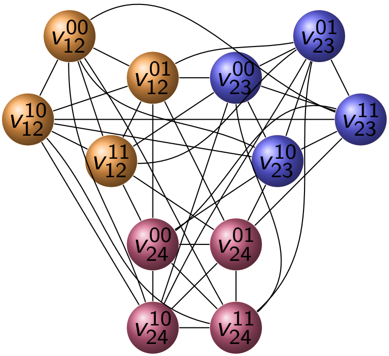
Reparameterizations and pruning

A binary edge potential can always be reparameterized (shifts score between potentials s.t. the total is unchanged; equivalent to soft arc consistency) so as to leave just one non-zero term, e.g.

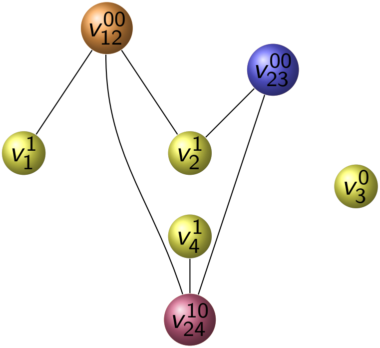
$$\underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{\substack{\text{original potential} \\ \psi_{ij}(x_i,x_j)}} = \underbrace{\begin{pmatrix} a+d-b-c & 0 \\ 0 & 0 \end{pmatrix}}_{\substack{\text{modified edge potential} \\ \psi'_{ij}(x_i,x_j)}} + \underbrace{\begin{pmatrix} c-d & 0 \\ c-d & 0 \end{pmatrix} + \begin{pmatrix} b-d & b-d \\ 0 & 0 \end{pmatrix}}_{\substack{\text{new unary potentials} \\ \psi'_i(x_i) \qquad \psi'_j(x_j)}} + \underbrace{\begin{pmatrix} d & d \\ d & d \end{pmatrix}}_{\text{constant}}$$

- This can be very powerful, allows us after pruning to end up with just one NMRF node per edge potential (instead of four);
- Though this may introduce new NMRF nodes for the unary terms.
- To show perfect, this seems very helpful and had been always used.
- In this work, we consider all reparameterizations: we show it can be good instead for some edges to keep all edge nodes and 'absorb' incident unary nodes.

Example: reparameterizing and pruning the earlier NMRF



Initial NMRF



After reparameterizing and pruning

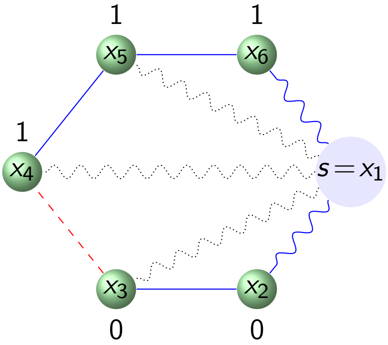
reparameterized s.t. all edges get one node

introduces new unary/singleton nodes

Example: application to a frustrated cycle

In the paper, we show constructively how MAP inference may be performed efficiently for any model composed of (possibly many) almost balanced blocks.

Blue edges are attractive, dashed red are repulsive. Straight edges are reparameterized s.t. they lead to one node in the pruned NMRF, wiggly edges may have all 4 possible nodes. Gray edges are ‘phantom edges’ introduced to absorb nodes from unary/singleton potentials. The special vertex s was chosen as x_1, removing this renders the remaining graph balanced (in fact acyclic in this example). Marks are shown next to their vertices for the two partitions in the balanced portion of the model. See paper for details.



Conclusion

- MAP inference is equivalent here to (soft) VCSP.
- The NMRF approach is a useful tool, equivalent to the complement of the microstructure of the dual of a VCSP.
- The method becomes more powerful by considering different reparameterizations (soft arc consistency) and pruning.
- Here we consider all possible reparameterizations and precisely characterize the limits of the approach for binary pairwise models using a signed graph topology (attractive/repulsive),
- Yielding a simple and interesting characterization - each block must be almost balanced - easy to check in polynomial time.

Thank you!

Contact: adrian.weller (at) eng.cam.ac.uk
Slides and related papers: http://mlg.eng.cam.ac.uk/adrian/

References

M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas. The strong perfect graph theorem. *Ann. Math*, 164:51–229, 2006.

M. Cooper and S. Živný. Hybrid tractability of valued constraint problems. *Artificial Intelligence*, 175(9):1555–1569, 2011.

A. El Mouelhi, P. Jégou, and C. Terrioux. Microstructures for CSPs with constraints of arbitrary arity. In *SARA*, 2013.

T. Jebara. MAP estimation, message passing, and perfect graphs. In *UAI*, 2009.

P. Jégou. Decomposition of domains based on the micro-structure of finite constraint-satisfaction problems. In *AAAI*, pages 731–736, 1993.

J. Larrosa and R. Dechter. On the dual representation of non-binary semiring-based CSPs. In *CP2000 workshop on soft constraints*, 2000.

A. Weller and T. Jebara. On MAP inference by MWSS on perfect graphs. In *UAI*, 2013.

Reduction to MWSS on an NMRF

Idea: A MAP configuration has
$\max_x \sum_c \psi_c(x_c) = \sum_c \max_{x_c} \psi_c(x_c)$ s.t. all the x_c are consistent,
consistency will be enforced by requiring a stable set.

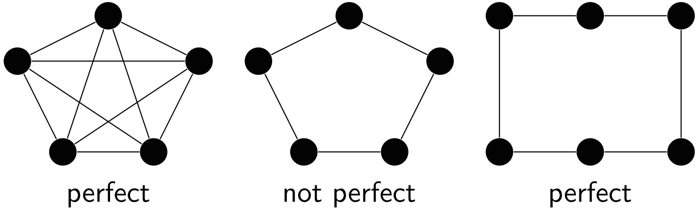
Given a model with potentials $\{\psi_c\}$ over variable sets $\{c\}$, construct a *nand Markov random field* (NMRF, Jebara, 2009) N, defined as follows:

- A weighted graph $N(V_N, E_N, w)$ with vertices V_N, edges E_N and a weight function $w : V_N \to \mathbb{Q}_{\geq 0}$.
- Each c of the original model maps to a clique in N. This contains one node for each possible configuration x_c, with all these nodes pairwise adjacent in N.
- Nodes in N are adjacent iff they have inconsistent settings for any variable X_i.
- Nonnegative weights of each node in N are set as $\psi_c(x_c) - \min_{x_c} \psi_c(x_c)$, hence the minimum weight is zero which facilitates *pruning*.

Perfect graphs

Berge defined perfect graphs in 1960: $\chi(H) = \omega(H) \quad \forall$ induced subgraphs $H \leq G$. The Strong Perfect Graph Theorem (Chudnovsky et al., 2006) yields an alternative definition:

- A graph is *perfect* iff it contains no *odd hole* or *odd antihole*.
- An *odd hole* is an induced subgraph which is a (chordless) odd cycle of length ≥ 4. An *antihole* is the complement of a hole (each edge of antihole is present iff not present in hole).



There is a rich literature on perfect graphs, e.g. pasting any 2 perfect graphs on a common clique yields a larger perfect graph.