Graphical Model Inference with Perfect Graphs

Tony Jebara
Columbia University

July 25, 2013

joint work with Adrian Weller

Graphical models and Markov random fields

- We depict a graphical model G as a bipartite factor graph with *variable* vertices $X = \{x_1, \ldots, x_n\}$ and square *factor* vertices $\Phi = \{\phi_1, \ldots, \phi_l\}$
- The x_i are discrete variables. The ϕ_i are positive functions.
- This compactly represents $p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c \in W} \phi_c(X_c)$ where X_c are variables that neighbor factor c

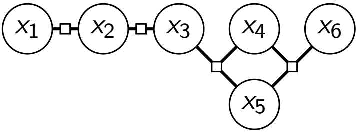
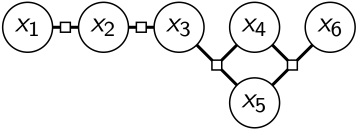


Figure: $p(X) = \frac{1}{Z}\phi_{1,2}(x_1, x_2)\phi_{2,3}(x_2, x_3)\phi_{3,4,5}(x_3, x_4, x_5)\phi_{4,5,6}(x_4, x_5, x_6)$.

Graphical models and Markov random fields



- Use marginal or maximum a posteriori (MAP) inference
  - Marginal inference: $p(x_i) = \sum_{X \setminus x_i} p(X)$
  - MAP inference: x_i^* where $p(X^*) \geq p(X)$
- In general:
  - Both are NP-hard (Cooper 1990, Shimony 1994)
  - Both are hard to approximate (Dagum 1993, Abdelbar 1998)
- But, on acyclic graphical models both are easy (Pearl 1988)

Belief propagation for tree inference

- Acyclic models are efficiently solvable by belief propagation
- Marginal inference via the sum-product:
  - Send messages from variable v to factor u

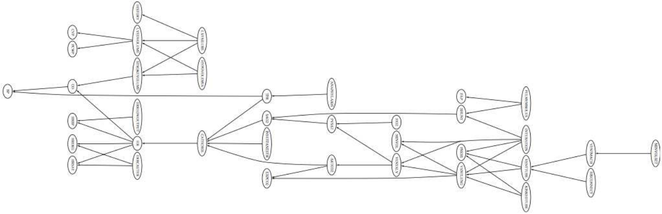
$$\mu_{v \to u}(x_v) = \prod_{u^* \in \mathrm{Ne}(v) \setminus \{u\}} \mu_{u^* \to v}(x_v)$$

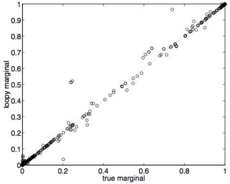
  - Send messages from factor u to variable v

$$\mu_{u \to v}(x_v) = \sum_{X'_u : x'_v = x_v} \phi_u(X'_u) \prod_{v^* \in \mathrm{Ne}(u) \setminus \{v\}} \mu_{v^* \to u}(x'_{v*})$$

  - Efficiently converge to $p(X_u) \propto \phi_u(X_u) \prod_{v \in \mathrm{Ne}(u)} \mu_{v \to u}(x_u)$
- MAP inference via max-product: swap $\sum_{X'_u}$ with $\max_{X'_u}$
- Informally, propagation also does well on loopy models!

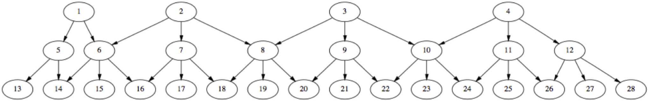
Loopy sum-product belief propagation

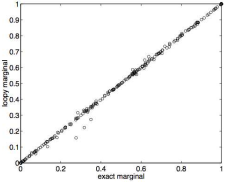




Alarm Network and Results

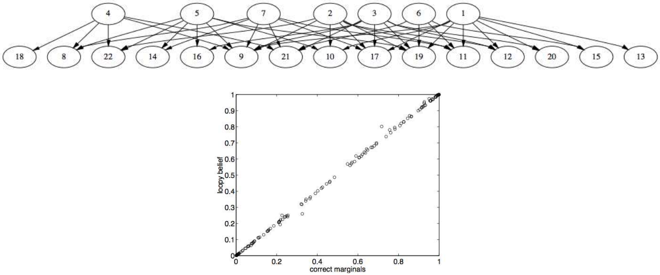
Loopy sum-product belief propagation





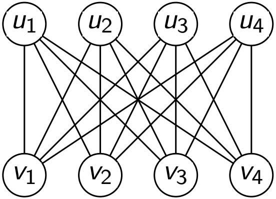
Pyramid Network and Results

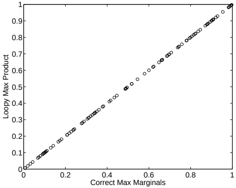
Loopy sum-product belief propagation



QMR-DT Network and Results

Loopy max-product belief propagation





Bipartite Matching Network and Results

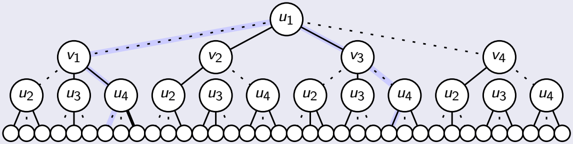
Loopy max-product

Theorem (HJ 2007)

Max product solves generalized bipartite matching MAP in $O(n^3)$ time.

Proof.

Using unwrapped tree T of depth $\Omega(n)$, we show that maximizing belief at root of T is equivalent to maximizing belief at corresponding node in original graphical model.



□

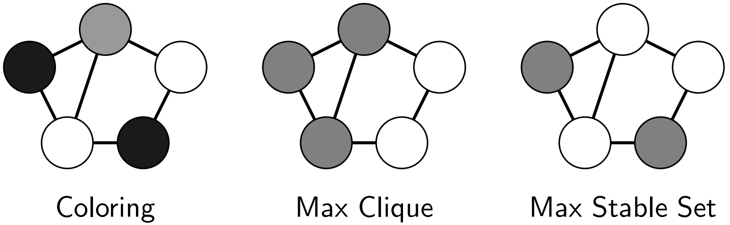
So some loopy graphical models might admit efficient inference...
Perfect graph theory characterizes the set of efficient loopy models!

What is a perfect graph?

Figure: Claude Berge

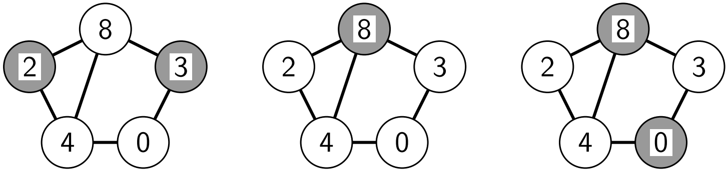
- In 1960, Berge introduced perfect graphs as
  - G perfect iff $\forall$induced subgraphs H, $\chi(H) = \omega(H)$
- Stated *Strong Perfect Graph Conjecture*, open for 50 years
- Many NP-hard problems solvable in polynomial time for perfect graphs (Grötschel Lovász Schrijver 1984)
  - Graph coloring
  - Maximum clique
  - Maximum stable set

Efficient problems on perfect graphs



- Coloring: color nodes with fewest colors such that no adjacent nodes have the same color
- Max Clique: largest set of nodes, all pairwise adjacent
- Max Stable Set: largest set of nodes, none pairwise adjacent

Efficient problems on weighted perfect graphs

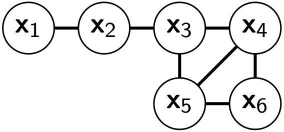


Stable set MWSS MMWSS

- Stable set: no two vertices adjacent
- Max Weight Stable Set (MWSS): stable set with max weight
- Maximal MWSS (MMWSS): MWSS with max cardinality
  (includes as many 0 weight nodes as possible)

MWSS via linear programming

$$\max_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\geq\mathbf{0}} \mathbf{f}^\top\mathbf{x} \text{ s.t. } \mathbf{A}\mathbf{x} \leq \mathbf{1}$$



$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

- $\mathbf{A} \in \mathbb{R}^{m\times n}$ is vertex versus maximal cliques incidence matrix
- $\mathbf{f} \in \mathbb{R}^n$ is vector of weights
- For perfect graphs, LP is binary and finds MWSS in $\mathcal{O}(\sqrt{m}n^3)$
- Note m is number of cliques in graph (may be exponential)

MWSS via semi-definite programming

$$\vartheta = \max_{\mathbf{M} \succeq \mathbf{0}} \sum_{ij} \sqrt{\mathbf{f}_i \mathbf{f}_j} \mathbf{M}_{ij} \text{ s.t. } \sum_{i} \mathbf{M}_{ii} = 1, \; \mathbf{M}_{ij} = 0 \, \forall (i,j) \in E$$

- This is known as the Lovász theta-function
- Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be the maximizer of $\vartheta_{\mathcal{F}}(\mathcal{G})$
- Let ϑ be the recovered total weight of the MWSS.
- Under mild assumptions, get $\mathbf{x}^* = \mathrm{round}(\vartheta \mathbf{M} \mathbf{1})$
- For perfect graphs, find MWSS in $\tilde{\mathcal{O}}(n^5)$

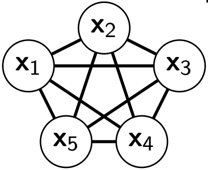
Other perfect graph theorems



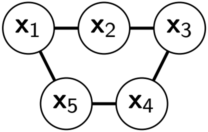

Theorem (Strong Perfect Graph Theorem, Chudnovsky et al 2006)

G perfect $\Leftrightarrow$ G contains no odd hole or antihole

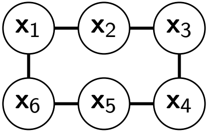
- Hole: an induced subgraph which is a (chordless) cycle of length at least 4. An odd hole has odd cycle length.
- Antihole: the complement of a hole



Perfect



Not Perfect

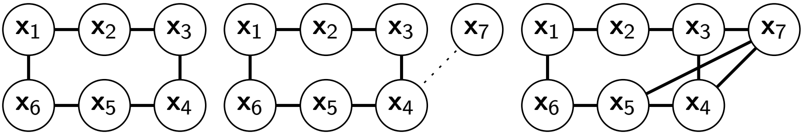


Perfect

Other perfect graph theorems

Lemma (Replication, Lovász 1972)

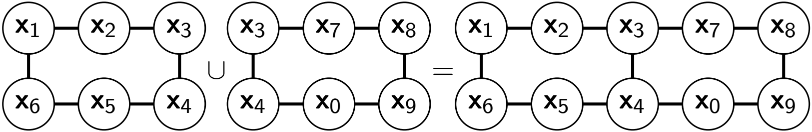
Let $\mathcal{G}$ be a perfect graph and let $v \in V(\mathcal{G})$. Define a graph $\mathcal{G}'$ by adding a new vertex v' and joining it to v and all the neighbors of v. Then $\mathcal{G}'$ is perfect.



Other perfect graph theorems

Lemma (Pasting on a Clique, Gallai 1962)

Let $\mathcal{G}$ be a perfect graph and let $\mathcal{G}'$ be a perfect graph. If $\mathcal{G} \cap \mathcal{G}'$ is a clique (clique cutset), then $\mathcal{G} \cup \mathcal{G}'$ is a perfect graph.



Our plan: reduce NP-hard inference to MWSS

- Reduce MAP to MWSS on weighted graph
- If reduction produces a perfect graph, inference is efficient
- Proves efficiency of MAP on
  - Acyclic models
  - Bipartite matching models
  - Associative models
  - Non frustrated models (new)
  - Slightly frustrated models (new)
- Reduce Bethe marginal inference to MWSS on weighted graph
- Proves efficiency of Bethe marginals on
  - Acyclic models
  - Associative models (new)

MAP inference

MAP inference: $x^* = \arg\max_x \sum_{c \in C} \psi_c(x_c)$

Where

- A set of n variables $V = \{x_1, \ldots, x_n\}$ together with (log) potential functions over subsets c of V, $\Psi = \{\psi_c : c \in C \subseteq \mathcal{P}(V)\}$
- Write $x = (x_1, \ldots, x_n)$ for one particular complete configuration, x_c for a configuration just of the variables in c
- A potential function ψ_c maps each possible setting x_c of its variables c to a real number $\psi_c(x_c)$
- $p(x) = \frac{1}{Z} \exp\left(\sum_{c \in C} \psi_c(x_c)\right)$

Assume

- n finite, $\psi_c(x_c)$ finite $\forall c, x_c$ so $p(x) > 0 \; \forall x$
- All x_i take values in a finite, discrete set

Reduction: MRF $M \longrightarrow$ NMRF N

Given an MRF model $M(V, \Psi)$, construct a *nand Markov random field (NMRF) N*:

- Weighted graph $N(V_N, E_N, w)$ with vertices V_N, edges E_N and weight function $w : V_N \longrightarrow \mathbb{R}_{\geq 0}$
- Each $c \in C$ from M maps to a *clique group* of N with one node for each configuration x_c, all pairwise adjacent
- Nodes are adjacent iff inconsistent settings for any variable X_i
- Weights of each node in N set as $\psi_c(x_c) - \min_{x_c} \psi_c(x_c)$

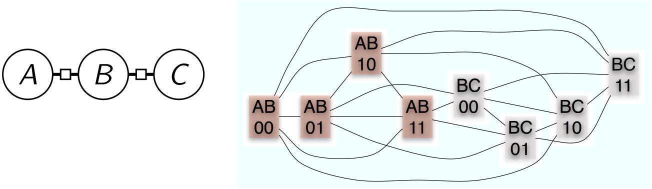
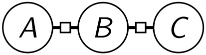
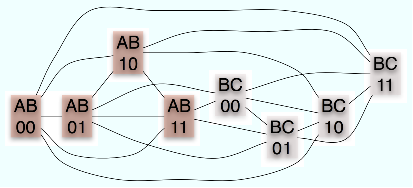


Figure: MRF M with binary variables (left) and NMRF N (right).

Reduction: MRF $M \longrightarrow$ NMRF N





MAP inference: identify $x^* = \arg\max_x \sum_{c \in C} \psi_c(x_c)$

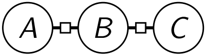
Lemma (J 2009)

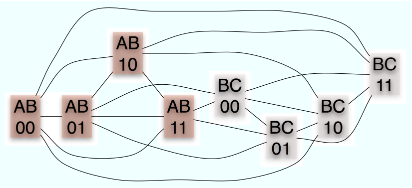
A MMWSS of the NMRF finds a MAP solution

Proof.

Sketch: *MAP selects, for each ψ_c, one configuration of x_c which must be globally consistent with all other choices, so as to max the total weight. This is exactly what MMWSS does.* □

NMRFs: a small example





AB clique group

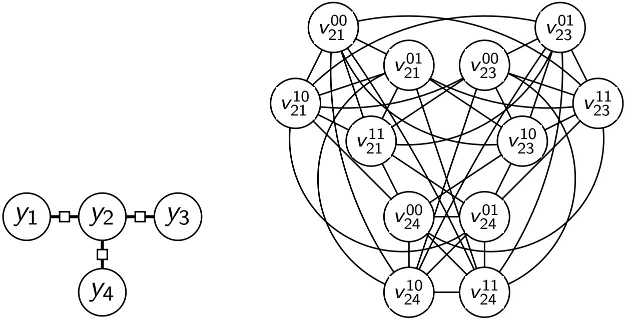
ψ/weight	A	B
5/0	0	0
5/0	1	0
7/2	0	1
8/3	1	1

BC clique group

B	C	ψ/weight
1	0	1/0
1	1	2/1
0	0	6/5
0	1	1/0

Figure: Graphical model's ψ values and final weights in NMRF

NMRF from a tree graphical model



(a) MRF (b) NMRF

Figure: Reducing a tree model

NMRF from a tree graphical model is perfect

Theorem (J 2009)

Let G be a tree, the NMRF $\mathcal{G}$ obtained from G is a perfect graph.

Proof.

First prove perfection for a star graph with internal node v with $|v|$ configurations. First obtain $\mathcal{G}$ for the star graph by only creating one configuration for non internal nodes. The resulting graph is a complete $|v|$-partite graph which is perfect. Introduce additional configurations for non-internal nodes one at a time using the replication lemma. The resulting $\mathcal{G}_{star}$ is perfect. Obtain a tree by induction. Add two stars $\mathcal{G}_{star}$ and $\mathcal{G}_{star'}$. The intersection is a fully connected clique (clique cutset) so by (Gallai 1962), the resulting graph is perfect. Continue gluing stars until full tree G is formed. □

Reparameterization and pruning

Lemma (WJ 2013)

To find a MMWSS, it is sufficient to prune any 0 weight nodes, solve MWSS on the remaining graph, then greedily reintroduce 0 weight nodes while maintaining stability.

A *reparameterization* is a transformation

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

Does not change the MAP solution but can simplify the NMRF

Lemma (WJ 2013)

MAP inference is tractable provided $\exists$ an efficient reparameterization s.t. we obtain a perfect pruned *NMRF*

Reparameterization and pruning

A *reparameterization* is a transformation

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

Does not change the MAP solution but can simplify the NMRF

One particular type of reparameterization

- *Singleton transformation*, i.e. change in one or more ψ functions for a single variable, corresponding changes simplify a higher order term
- Very helpful for pruning nodes from edge (or higher order) clique groups
- To allow this in general for arbitrary ψ, we assume all NMRFs have all singleton nodes

Singleton transformation: a small example

- Binary variables x_1 and x_2 connected by an edge

$\psi_1(x_1)$

x_1	$\psi_1(x_1)$
0	2
1	4

$\psi_{12}(x_1, x_2)$

$x_1 \backslash x_2$	0	1
0	1	3
1	5	2

$\psi_2(x_2)$

x_2	0	1
$\psi_2(x_2)$	1	-1

$\psi'_1(x_1)$

x_1	$\psi'_1(x_1)$
0	2
1	3

$\psi'_{12}(x_1, x_2)$

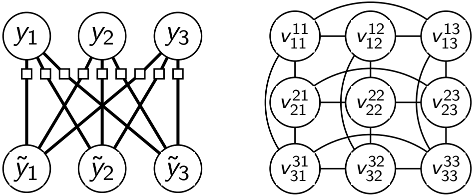
$x_1 \backslash x_2$	0	1
0	0	0
1	5	0

$\psi'_2(x_2)$

x_2	0	1
$\psi'_2(x_2)$	2	2

- After reparameterization, pruned NMRF for $\{\psi'\}$ has just one node $(x_1 = 1, x_2 = 0)$ for the edge clique group.

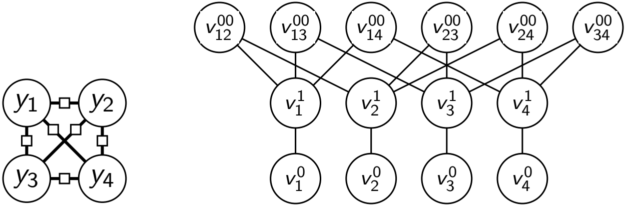
NMRF from a matching graphical model is perfect



(a) MRF (b) pruned NMRF

Figure: Reducing a matching model

NMRF from an associative graphical model is perfect



(a) MRF (b) pruned NMRF

Figure: Reducing an associative binary pairwise model

General decomposition theorem

Theorem (General MRF Decomposition, WJ 2013)

If $MRF_A(V_A, \Psi_A)$ and $MRF_B(V_B, \Psi_B)$ both map to perfect NMRFs N_A and N_B, and have exactly one variable s in common, i.e. $V_A \cap V_B = \{s\}$, then the combined $MRF'(V_A \cup V_B, \Psi_A \cup \Psi_B)$ maps to an NMRF N' which is also perfect. The converse is true by the definition of perfect graphs.

(Neat proof using paste perfect graphs on a star-cutset)

Corollary (Block Decomposition for pairwise models)

A pairwise MRF maps to a perfect NMRF for all valid ψ iff each of its blocks maps to a perfect NMRF.

where a *block* is a maximal 2-connected subgraph

Corollary: neatly confirms earlier result for a tree

Review of terms for connectivity, blocks for a graph G

- G is *connected* if there is a path connecting any two vertices
- A *cut vertex* of a connected graph G is a vertex $v \in V$ such that deleting v disconnects G
- G is *2-connected* (*biconnected*) if connected, no cut vertex
- A *block* is a maximal connected subgraph with no cut vertex

Every block is either K_2 (two vertices joined by an edge) or a maximal 2-connected subgraph containing a cycle. Different blocks of G overlap on at most one vertex, which must be a cut vertex. Hence G can be written as the union of its blocks with every edge in exactly one block. These blocks are connected without cycles in the *block tree* for G.

Example of block decomposition for a small graph

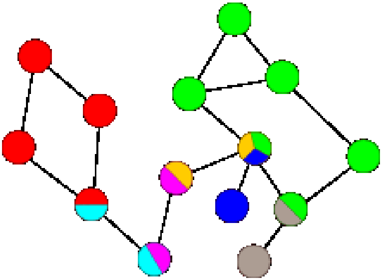


Figure: Each color corresponds to a block. Multi-colored vertices are cut vertices, hence belong to multiple blocks.

Which binary pairwise MRFs M map to perfect NMRFs N?

- Characterize MRFs based only on the sign of its edges (attractive/repulsive), otherwise free to choose any valid ψ

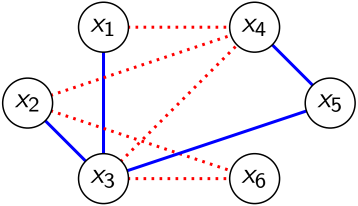


Figure: A signed graph, solid (dashed) edges are attractive (repulsive)

Results:

- By decomposition result, need only consider 2-connected blocks
- By Strong Perfect Graph Theorem, look for potential odd holes or antiholes in N, consider which structures in M could cause them

Which binary pairwise MRFs M map to perfect NMRFs N?

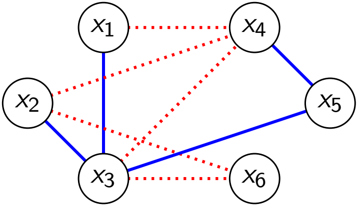


Figure: A signed graph, solid (dashed) edges are attractive (repulsive)

Remark: a hole of size 5 is equivalent to an antihole of size 5

Additional Results:

- Lemma: never get an antihole of size ≥ 7
- Hence, need only look for potential odd holes
- Lemma: an odd hole in N can only derive from a cycle in M with an odd number of repulsive edges, i.e. a frustrated cycle

Which binary pairwise MRFs M map to perfect NMRFs N?

Consider cases

- No frustrated cycle in M: what we call a B_R structure
- Frustrated cycle with > 3 edges: cannot avoid an odd hole in N
- Frustrated cycle with exactly 3 edges
    - 1 repulsive edge: to avoid odd holes must have U_n structure
    - 3 repulsive edges: to avoid odd holes must have $T_{m,n}$ structure

Theorem

A binary pairwise MRF maps to a perfect pruned NMRF for all valid ψ_c iff each of its blocks (using all edges) has the form B_R, $T_{m,n}$ or U_n.

Example of a B_R structure

Lemma (Harary 53)

The following are equivalent for a 2-connected signed graph G:

1. *G contains no frustrated cycle (B_R)*
2. *G is flippable to fully attractive*

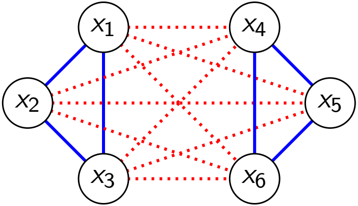


Figure: A B_R structure is 2-connected and contains no frustrated cycle. Solid (dashed) edges are attractive (repulsive). Deleting any edges maintains the B_R property.

Examples of $T_{m,n}$ and U_n structures

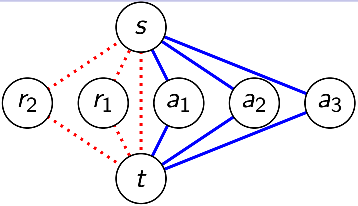


Figure: A $T_{m,n}$ structure with $m = 2$ and $n = 3$. Note triangle with 3 repulsive edges. Solid (dashed) edges are attractive (repulsive).

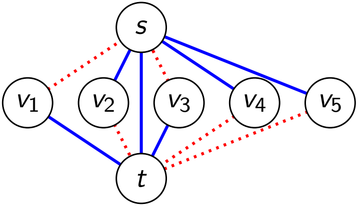


Figure: A U_n structure with $n = 5$. Note triangle with 1 repulsive edge. Solid (dashed) edges are attractive (repulsive).

Example combining different blocks

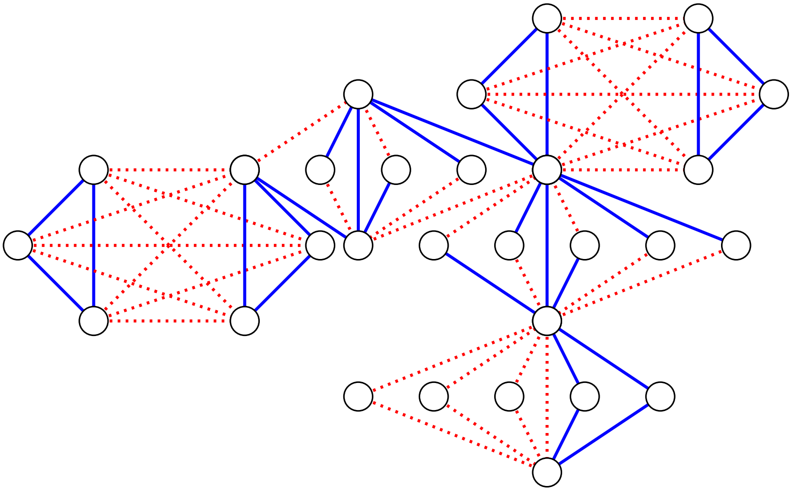


Figure: Binary pairwise MRF, provably tractable with perfect pruned NMRF due to block tree decomposition into B_r, $T_{m,n}$ and U_n structures.

Summary: reduce NP-hard inference to MWSS

- Reduce MAP to MWSS on weighted graph
- If reduction produces a perfect graph, inference is efficient
- Proves efficiency of MAP on
  - Acyclic models
  - Bipartite matching models
  - Associative models
  - Models without frustrated cycles (new)
  - Slightly frustrated models with frustrated 3-cycles (new)
- Reduce Bethe marginal inference to MWSS on weighted graph
- Proves efficiency of Bethe marginals on
  - Acyclic models
  - Associative models (new)

Thank you!

- B. Huang and T. Jebara. "Loopy Belief Propagation for Bipartite Maximum Weight b-Matching". Artificial Intelligence and Statistics (AISTATs), 2007.
- T. Jebara. "MAP Estimation, Message Passing, and Perfect Graphs". Uncertainty in Artificial Intelligence (UAI), 2009.
- T. Jebara. "Perfect Graphs and Graphical Modeling". Tractability: Practical Approaches to Hard Problems, Cambridge University Press, 2013.
- A. Weller and T. Jebara. "Bethe Bounds and Approximating the Global Optimum". Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATs), 2013.
- A. Weller and T. Jebara. "On MAP Inference by MWSS on Perfect Graphs". Uncertainty in Artificial Intelligence (UAI), 2013.

- Funded by "RI: Small: Learning and Inference with Perfect Graphs". NSF IIS-1117631, T. Jebara and M. Chudnovsky.