Dynamic Planar-Cuts: Efficient Computation of Min-Marginals for Outer-Planar Models

Dhruv Batra[1]
[1]Carnegie Mellon University
www.ece.cmu.edu/~dbatra

Tsuhan Chen[2,1]
[2]Cornell University
tsuhan@ece.cornell.edu

Abstract

This paper deals with *Dynamic* MAP Inference, where the goal is to solve an instance of the MAP problem given that we have already solved a *related* instance of the problem. We propose an algorithm for Dynamic MAP Inference in planar Ising models, called Dynamic Planar-Cuts. As an application of our proposed approach, we show that we can extend the MAP inference algorithm of Schraudolph and Kamenetsky [14] to efficiently compute min-marginals for all variables in the same time complexity as the MAP inference algorithm, which is an $O(n)$ speedup over a naïve approach.

1 Introduction

One of the classical examples of discrete optimization problems found in machine learning is that of maximum *a posteriori* (MAP) estimation in undirected graphical models. Specifically, we consider a set of discrete random variables $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$, where each variable X_i takes values in some label space $L = \{1, 2, \ldots, k\}$. Without loss of generality, we may assume a *pairwise* Markov Random Field (MRF), given by a graph $G = (\mathcal{V}, \mathcal{E})$, where each variable is represented by a node ($\mathcal{V} = \{1, 2, \ldots, n\}$), and the edge-set ($\mathcal{E}$) is the union of the Markov blanket of each X_i. The goal of MAP inference is to minimize a real-valued energy function associated with this graph, *i.e.*:

$$\begin{aligned} \mathcal{X}^* &= \operatorname*{argmin}_{\mathcal{X} \in \mathcal{L}} E(\mathcal{X}) && \text{(1a)} \\ &= \operatorname*{argmin}_{\mathcal{X} \in \mathcal{L}} \sum_{i \in \mathcal{V}} E_i(X_i) + \sum_{(i,j) \in \mathcal{E}} E_{ij}(X_i, X_j), && \text{(1b)} \end{aligned}$$

where the first term E_i is called the node/unary energy, and the second term E_{ij} is called the edge/pairwise energy.

In general, this problem is known to be NP-hard [15]. However, certain subclasses are known where exact efficient inference is possible. Trees and polytrees [13] were the first known efficient structures. For 2-class problems ($k = 2$), Kolmogorov and Zabih [11] showed that exact inference can be performed in polynomial time if the energy functions are sumodular. More recently, Schraudolph and Kamenetsky (SK) [14] showed that exact inference can be performed efficiently in planar Ising models, which are equivalent to general outer-planar models (*i.e.*, models where energy is given by (1b) and G is an outer-planar graph).

Contributions. This paper deals with "Dynamic" MAP Inference in planar Ising models, where the goal is to solve an instance of the problem given that we have already solved a *related* instance of the problem. Specifically, if we already minimized one such energy function, and only a few energy terms in Eqn. 1b change, we show how we can re-use the computations from the planar-cut algorithm of SK [14] to compute the new solution. As an application of our algorithm, we consider the problem of computing min-marginals (described in Sec. 3). We show that we can extend the MAP inference algorithm of SK [14] to efficiently compute *exact* min-marginals for all variables in the *same* time complexity as the MAP algorithm. Specifically, a naïve extension would compute min-marginals via $2n$-min-cuts in a constructed graph (discussed next). On the other hand, our algorithm performs 1-min-cut computation followed by n-update steps, each of which is significantly ($O(n)$) faster than computing a min-cut "from scratch".

Applications. The motivation for Dynamic MAP Inference may come from a variety of applications that have a sequence of related problems as input. We discuss a few from the field of computer vision. Kohli and colleagues applied dynamic inference for object-background segmentation in video [7] and human pose estimation and segmentation [6]. Similarly, min-marginals have been used for a variety of tasks. Glocker *et al*. [4] used min-marginals for optical-flow estimation. Kohli *et al*. [8] used min-marginals for a parameter learning scheme and as a confidence measure with interactive image segmentation.

Relations to previous work. We note that Kohli *et al*. [8] have also developed dynamic algorithms for s-t min-cut based MAP inference in submodular energy functions. While the high level goal of our paper is similar to theirs, the techniques developed in these two papers are completely different. Their min-cut algorithm is based on max-flow computations, while ours is based on perfect matchings in expanded planar-dual graphs, and unfortunately, none of their techniques are applicable here.

Organization. The rest of this paper is organized as follows: Section 2 recalls the construction of SK [14] to show how MAP inference in planar Ising models can be formulated as a min-cut problem in a planar graph; Section 3 presents our proposed approach of Dynamic Planar-Cuts; Section 4 presents a timing comparison of Dynamic Planar-Cuts vs. naïve computation of min-marginals; Finally, Section 5 concludes with discussions.

2 MAP as Min-Cut in a Planar Graph

We now recall the construction of Schraudolph and Kamenetsky [14] for planar Ising models.

Planar Ising Models. Ising models have a rich history in statistical physics [1] and are defined by an energy function (over boolean variables) with the following form:

$$\tilde{E}(\mathcal{X}) = \sum_{(i,j)\in\mathcal{E}} [\![X_i \neq X_j]\!]\, \tilde{E}_{ij}, \tag{2}$$

where $[\![\cdot]\!]$ is the indicator function. While this energy ($\tilde{E}$) seems much more restrictive (no node terms, symmetric edge terms) than the general one in (1b), the following theorem (from [14]) shows that they are in fact equivalent.

Theorem 1 *Every energy function of the form (1b) over n boolean variables is equivalent to an Ising energy function (2) over $n+1$ boolean variables with the additional variable held constant.*

As a consequence of this theorem, we can establish the equivalence between *planar* Ising models and *outer-planar* general models.

Corollary 1 *Every* outer-planar *energy function of the form (1b) over n boolean variables is equivalent to a* planar *Ising energy function (2) over $n+1$ boolean variables with the additional variable held constant.*

Definition. Outer-planar graphs are a sub-class of planar graphs – they allow a planar embedding in which all nodes lie on a common external unbounded face. An alternate definition is more helpful for our purpose. A graph $G = (\mathcal{V}, \mathcal{E})$ is outer-planar iff the modified graph $G_c = (\mathcal{V} \cup \{o\}, \mathcal{E} \cup \{(o,i) : i \in \mathcal{V}\})$ formed by adding an extra node v and connecting it to all nodes in G is planar.

From this point on our discussion assumes that we are given an outer-planar model ($G = (\mathcal{V}, \mathcal{E})$) with energy function given by (1b).

Construction 1. It can be shown [14] that MAP inference (1a) with boolean variables is equivalent to finding a min-cut in a graph ($G_c = (\mathcal{V} \cup \{o\}, \mathcal{E} \cup \{(o,i) : i \in \mathcal{V}\})$) constructed as follows:

1. For each node energy $E_i(X_i)$, we set the edge weight on (o,i) as $w_{(o,i)} = E_i(1) - E_i(0)$.

2. For each edge energy function $E_{ij}(X_i, X_j)$, add the three edges weights:

$$w_{(o,i)} \mathrel{+}= \frac{1}{2}\left[E_{ij}(1,0) + E_{ij}(1,1) - E_{ij}(0,1) - E_{ij}(0,0)\right] \tag{3a}$$

$$w_{(o,j)} \mathrel{+}= \frac{1}{2}\left[E_{ij}(0,1) + E_{ij}(1,1) - E_{ij}(0,0) - E_{ij}(1,0)\right] \tag{3b}$$

$$w_{(i,j)} = \frac{1}{2}\left[E_{ij}(1,0) + E_{ij}(0,1) - E_{ij}(0,0) - E_{ij}(1,1)\right] \tag{3c}$$

Claim 1 *There exists a bijection between labellings and cuts. Every labelling $\mathcal{X} \in \{0,1\}^n$ induces a cut $\mathcal{C}$ in G_c s.t. $\mathcal{C} = \{(i,j) \in \mathcal{E}_c : X_i \neq X_j\}$ (where node o is held at label 0,* i.e. *$X_o = 0$).*

Claim 2 *The energy of a labelling $\mathcal{X}$ is equal to the cost of its induced cut $\mathcal{C}$ (plus a constant): $E(\mathcal{X}) = w(\mathcal{C}) + E(\mathbf{0})$ (where $\mathbf{0}$ is a the zero labelling, $X_u = 0\,\forall u$).*

Proof See [14].

Thus, this construction allows us to find the MAP state by computing the min-cut in this constructed graph G_c. We note that this construction is related to the one proposed by Kolmogorov and Zabih (KZ) [11], however there are two key differences. First, this is an *undirected* graph characterization while KZ presented a *directed* graph representation of energy functions. Second (and more importantly), this construction makes no assumption of submodularity. This implies that the edge weights w_{ij} may be negative and thus max-flow based algorithms cannot be used to compute this min-cut. This inapplicability of max-flow algorithms is not just for lack of some clever trick. There is a more fundamental barrier at play here. The general min-cut problem with both positive and negative edge-weights is just as hard as the max-cut problem, which is NP-complete. Fortunately, planar graphs are a special subclass for which max-cut (and thus min-cut) can be computed in polynomial time via max-weight perfect matching in an expanded planar dual graph. Due to lack of space, only a brief description, necessary for understanding our contribution, is provided below. For details the reader is referred to [5, 12, 14].

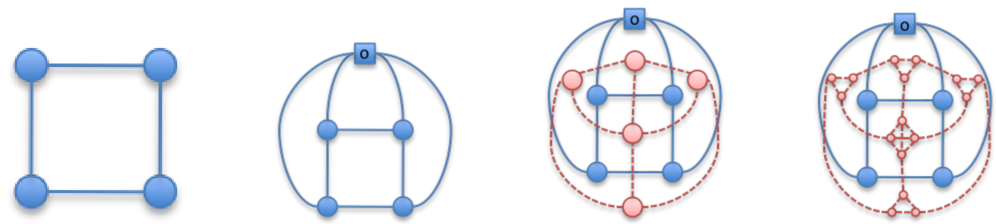
2.1 Min-cut via Perfect-matching in Expanded Dual

A planar dual graph $G^* = (\mathcal{V}^*, \mathcal{E}^*)$ is constructed by transforming faces in G_c to nodes in G^*. Specifically, G^* has a vertex corresponding to each face in G_c and two vertices in G^* are joined by an edge if their corresponding faces share an edge. An *expanded* dual graph $G_{ex} = (\mathcal{V}_{ex}, \mathcal{E}_{ex})$ is constructed by replacing each node v in G^* by a $d(v)$-clique (where $d(v)$ is the degree of v). There is a natural 1-1 correspondence between the edge-sets of G_c and G^* ($\mathcal{E}_c \Leftrightarrow \mathcal{E}^*$), which leads to an injective mapping between $\mathcal{E}_c$ and $\mathcal{E}_{ex}$ (*i.e.*, nothing maps onto the edges in $\mathcal{E}_{ex}$ created due to the introduction of cliques). Let's call this injective mapping $\mathcal{F}$. We copy the edge-weights from G_c over to their corresponding edges in G_{ex} (*i.e.*, $w_{\mathcal{F}(e)} = w_e$, $\forall e \in \mathcal{E}_c$) and clique edges have zero weights. An example is shown in Fig. 1.

Now, it is known [12, 14] that we can establish the following correspondence between perfect matchings in G_{ex} and cuts in G_c:

Theorem 2 *Every perfect matching $\mathcal{P}$ in G_{ex} of weight M corresponds to cut $\mathcal{C}$ in $G_c = (\mathcal{V}_c, \mathcal{E}_c)$ of weight $\sum_{(i,j)\in\mathcal{E}_c} w_{(i,j)} - M$. In addition, an edge e of G_c is part of $\mathcal{C}$, iff its corresponding edge f in G_{ex} is* not *part of $\mathcal{P}$,* i.e.*, $e \in \mathcal{C} \Longleftrightarrow \mathcal{F}(e) \notin \mathcal{P} \quad (\forall e \in \mathcal{E}_c)$.*

Thus the min-cut in G_c can be found by looking for the max-weight perfect matching in G_{ex}. Max-weight perfect matching is a well studied problem, first solved by the Blossom algorithm of Edmonds [3]. It has a long history of improvements and efficient implementations [2, 10], and the best known algorithm is $O(n(m + \log n))$. We use the BlossomV implementation of Kolmogorov [10], which although is asymptotically slower $O(n^2 m)$, but very efficient in practice and able to handle hundreds of thousands of nodes. For the purpose of understanding this paper, it is sufficient to recall that the "vanilla" Blossom algorithm of Edmonds [3] is an iterative algorithm that maintains a matching (of cardinality $< n/2$) and at each step updates this matching by finding an augmenting path so that the cardinality of this matching increases by 1 (till it reaches $n/2$). The fact that Blossom requires $n/2$ augmentations will be useful in understanding our contribution.



(a) G (b) G_c (c) G^* (d) G_{ex}

Figure 1: MAP in Outer-Planar Models: (a) shows an outer-planar graph G; (b) shows the result of Construction 1: G_c; (c) shows the planar dual graph G^* overlaid on G_c. Notice that there is a natural 1-1 correspondence between the edges sets of G_c and G^*; (d) shows the expanded planar dual G_{ex} overlaid on G_c. Each node v in G^* has been replaced by a $d(v)$-clique. A perfect matching in G_{ex} corresponds to a cut in G_c, which corresponds to a labelling in G.

3 Dynamic Planar-Cuts for Min-marginals

We now present our proposed approach of Dynamic Planar-Cuts. As we discussed earlier, a lot of applications have a natural dynamic nature, *i.e.*, they have a sequence of related inputs, and it makes sense to re-use computation from one solved instance to solve the next. To explain our algorithm in detail, we focus on the problem of computing min-marginals. However, we note that the ideas described in this paper are general enough to applied to any of the applications mentioned before.

Min-marginals. A min-marginal is a function that describes the energy function under certain constraints. Specifically, it is the minimum value of the energy function when the state for a single node is specified. Following the notation of Kolmogorov [9], we write min-marginals as:

$$\Phi_i(x_i) = \min_{\mathcal{X}, X_i = x_i} E(\mathcal{X}) \quad \forall i \in \mathcal{V}, x_i \in \mathcal{L} \tag{4}$$

We note that the max-marginals produced as messages in max-product belief propagation algorithms are related to min-marginals as:

$$\mu_i(x_i) = \max_{\mathcal{X}, X_i = x_i} \Pr(\mathcal{X}) = \max_{\mathcal{X}, X_i = x_i} \frac{1}{\mathcal{Z}} \exp\left(-E(\mathcal{X})\right) = \frac{1}{\mathcal{Z}} \exp(-\Phi_i(x_i)) \tag{5}$$

So how would we compute these min-marginals for boolean outer-planar models? First, we note that these min-marginals may be represented by a $2 \times n$ matrix $\mathbf{\Phi} = [\Phi_1 \; \Phi_2 \ldots \Phi_n]$, where each column represents the min-marginal function for a node. The naïve way to compute the elements of this matrix is to loop through nodes, *force* a node X_i to take a particular state x_i and compute the MAP state and energy using the construction described above. We note that a node X_i can be forced to take state 0(1) by setting the weight for (o, i) edge as $-\infty(+\infty)$. Clearly, this requires $2n$-min-cuts. To be fair, this naïve approach can be slightly improved to $(n+1)$-min-cuts by first computing the MAP energy (1-min-cut) and using that to fill n out of the $2n$ entries in this matrix. This as shown in Alg. 1.

Input: G_c: Output of Construction 1, and G_{ex}: Expanded planar dual graph.
Output: $\mathbf{\Phi}$: Min-marginals.
Find min-cut in G_c (via max-wt perfect matching in G_{ex}) to get MAP state $\mathcal{X}^*$ and energy ϵ.
for *i = 1 to n* **do**
  $\Phi_i(X_i^*) \longleftarrow \epsilon$;
  $G_c^{(i)} \longleftarrow G_c$; $G_{ex}^{(i)} \longleftarrow G_{ex}$;
  In $G_c^{(i)}$, set edge-weight $w_{(o,i)} \longleftarrow (2X_i^* - 1) \cdot (\infty)$; Update $G_{ex}^{(i)}$;
  Find min-cut in updated $G_c^{(i)}$ and compute energy (ϵ_i) corresponding to this cut;
  $\Phi_i(1 - X_i^*) \longleftarrow \epsilon_i$;
end

Algorithm 1: Naïve computation of min-marginals.

We note that this naïve algorithm (Alg. 1), repeatedly solves n very similar problems in a loop. In fact, there is a rich structure to these problems – each successive problem differs from the previous one in only 2 edge-weights ($w_{(o,i-1)}$ and $w_{(o,i)}$), and all other edge-weights are exactly the same.

Clearly, we should be able to exploit this structure, and solve each problem faster than solving "from scratch". Indeed, as we show next it is possible to solve all n of these problems in the same time as 1-min-cut ($O(n^2m)$), thus bringing the complexity of computing min-marginals to the same level as computing the MAP solution.

The key insight in achieving this is to notice that a solution to the problem at iteration $(i-1)$ isn't very far (in a combinatorial sense) from the solution to the next problem (iteration i). The trick though, is to look at the *right* entity. Specifically, let $\mathcal{C}^{(i)}$ be the min-cut in $G_c^{(i)}$ and $\mathcal{P}^{(i)}$ be the corresponding max-weight perfect-matching in $G_{ex}^{(i)}$ at iteration (i). Now, the two successive min-cuts $\mathcal{C}^{(i-1)}$ and $\mathcal{C}^{(i)}$ may be arbitrarily different, however the two perfect-matchings $\mathcal{P}^{(i-1)}$ and $\mathcal{P}^{(i)}$ are not too far away from each other, in the sense of augmenting paths.

We claim that an adjusted version of $\mathcal{P}^{(i-1)}$ is at most 2 augmenting operations away from $\mathcal{P}^{(i)}$. Specifically, $\mathcal{P}^{(i-1)}$ is also an $n/2$ sized matching in $G_{ex}^{(i)}$, however we need to adjust for the edge $\mathcal{F}(e)$, where $e = (o, i)$. If the updated weight w_e is $-\infty$, *i.e.*, e is forced to be in $\mathcal{C}^{(i)}$ and thus forced to not be in $\mathcal{P}^{(i)}$, and $\mathcal{P}^{(i-1)}$ contains $\mathcal{F}(e)$, then we can just delete $\mathcal{F}(e)$ from $\mathcal{P}^{(i-1)}$, resulting in an $n/2-1$ sized matching. Alternatively, if the updated weight w_e is $+\infty$, *i.e.*, $\mathcal{F}(e)$ forced to be in $\mathcal{P}^{(i)}$, and $\mathcal{P}^{(i-1)}$ does not contains $\mathcal{F}(e)$, we can just delete the two matching edges covering nodes o and i in $\mathcal{P}^{(i-1)}$, resulting in a matching of size $n/2-2$.

It is precisely the above observation that allows us to efficiently compute min-marginals by re-using the perfect matching solution from the previous iteration. At each iteration, we need to simply perform at most *two* augmentation steps, as opposed to $n/2$ augmentations required in solving from scratch, thus giving us a speedup by a factor of $O(n)$. Additionally, it can be shown that each augmentation step takes at most $O(nm)$ time (see for example, the discussion in Sec. 2.3 in [10]). Thus each update step in our algorithm takes $O(nm)$, and the total time for computing all min-marginal is $O(n^2m)$, which is precisely the complexity of a single perfect-matching/min-cut computation. Thus, using our approach, the extension from a single MAP solution to all min-marginals only adds a constant-factor overhead. This is summarized in Alg. 2.

```
Input: G_c: Output of Construction 1, and G_ex: Expanded planar dual graph.
Output: Φ: Min-marginals.
Find min-cut in G_c (max-wt perfect matching P in G_ex) to get MAP state X* and energy ε.
Set P^(0) ⟵ P;
for i = 1 to n do
    Φ_i(X*_i) ⟵ ε; G_c^(i) ⟵ G_c; G_ex^(i) ⟵ G_ex;
    In G_c^(i), set edge-weight w_(o,i) ⟵ (2X*_i − 1)·(∞); Update G_ex^(i);
    Adjust P^(i−1): if (X*_i == 0) then
        P^(i−1) ⟵ P^(i−1) − F((o,i));
    else
        P^(i−1) ⟵ P^(i−1) − {f_1, f_2 ∈ P^(i−1) : o ∈ f_1, i ∈ f_2};
    end
    Augment P^(i−1) to find P^(i) and compute energy (ε_i) corresponding to C^(i);
    Φ_i(1 − X*_i) ⟵ ε_i;
end
```

Algorithm 2: Dynamic Planar-Cuts.

4 Experiments

We now describe a timing comparison of the naïve approach (Alg. 1) vs. our approach of Dynamic Planar-Cuts (Alg. 2). We worked with a $2 \times n$ sized grid graph, which is guaranteed to be outer-planar. Following the setup of Kolmogorov [9], we created artificial energy functions for this graph by randomly sampling from Gaussians: $E_i(0), E_i(1) \sim \mathcal{N}(0,1)$. Pairwise energies were set as: $E_{ij}(0,0) = E_{ij}(1,1) = 0$; $E_{ij}(0,1), E_{ij}(1,0) \sim \mathcal{N}(0,1)$. Fig. 2 shows the time taken to compute the min-marginals for these energies. We can see that our approach is able to compute min-marginals significantly faster than the naïve approach. In fact, as the size of the graph increase the naïve approach becomes prohibitively slow. We point out that our approach is not an approximation, it computes *exact* min-marginals.

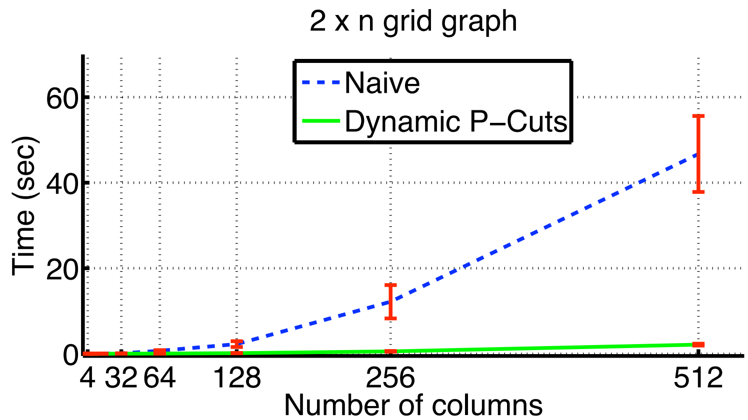


Figure 2: Time taken to compute min-marginals for a $2 \times n$ grid graph.

5 Discussions

To summarize, this paper proposed an algorithm for Dynamic MAP Inference in outer-planar models, where the goal is to solve an instance of the problem given that we have already solved a *related* instance of the problem. As an application of our Dynamic Planar-Cut algorithm, we showed that we can extend the MAP inference algorithm of SK [14] to efficiently and exactly compute all min-marginals, with a constant-factor overhead, which is $O(n)$ speedup over a naïve approach.

Extensions. Even though the details of our algorithm are discussed for computing min-marginals, we note that the ideas described in this paper are general enough to applied to any of the dynamic applications mentioned in the introduction. In addition, while our algorithm only shows how to compute node min-marginals, it can be easily extended to compute edge min-marginals, *i.e.*, minimum value of the energy function when the states for a *pair* of nodes is specified.

Acknowledgments. We thank Nic Schraudolph and Vladimir Kolmogorov for helpful discussions and making their implementations publicly available, without which this work would not be possible. We also thank Devi Parikh for helpful discussions.

References

[1] F. Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241–3253, 1982.

[2] W. Cook and A. Rohe. Computing Minimum-Weight Perfect Matchings. *Informs Journ. on Computing*, 11(2):138–148, 1999.

[3] J. Edmonds. Maximum matching and a polyhedron with (0, 1) vertices. In *J. of Res. at Nat. Bur. Stands.*, page 68, 1965.

[4] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Optical flow estimation with uncertainties through dynamic mrfs. In *CVPR*, 2008.

[5] P. W. Kasteleyn. Dimer statistics and phase transitions. *Journ. of Math. Phys.*, 4(2):287–293, 1963.

[6] P. Kohli, J. Rihan, M. Bray, and P. H. Torr. Simultaneous segmentation and pose estimation of humans using dynamic graph cuts. *IJCV*, 79(3):285–298, 2008.

[7] P. Kohli and P. H. S. Torr. Dynamic graph cuts for efficient inference in markov random fields. *PAMI*, 29(12):2079–2088, 2007.

[8] P. Kohli and P. H. S. Torr. Measuring uncertainty in graph cut solutions. *CVIU*, 112(1):30–38, 2008.

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.

[10] V. Kolmogorov. Blossom V: a new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation*, 1(1):43–67, 2009.

[11] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.

[12] F. Liers and G. Pardella. A simple max-cut algorithm for planar graphs. Technical report, Combinatorial Optimization in Physics (COPhy), 2008.

[13] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *AAAI*, 1982.

[14] N. N. Schraudolph and D. Kamenetsky. Efficient exact inference in planar ising models. In *NIPS*, 2008.

[15] S. E. Shimony. Finding maps for belief networks is np-hard. *Artificial Intelligence*, 68(2):399–410, August 1994.