# Approximating the Partition Function by Deleting and then Correcting for Model Edges (Extended Abstract)

**Arthur Choi and Adnan Darwiche**
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095
{aychoi,darwiche}@cs.ucla.edu

## Abstract

We propose an approach for approximating the partition function which is based on two steps: (1) computing the partition function of a simplified model which is obtained by deleting model edges, and (2) rectifying the result by applying an edge-by-edge correction. The approach leads to an intuitive framework in which one can trade-off the quality of an approximation with the complexity of computing it. It also includes the Bethe free energy approximation as a degenerate case. We develop the approach theoretically in this paper and provide a number of empirical results that reveal its practical utility.

## 1 Introduction

We presented in prior work an approach to approximate inference which is based on performing exact inference on a simplified model (Choi & Darwiche, 2006a, 2006b). We proposed obtaining the simplified model by deleting enough edges to render its treewidth manageable under the current computational resources. Interestingly enough, the approach subsumes iterative belief propagation (IBP) as a degenerate case, and provides an intuitive framework for capturing some of its generalizations (Choi & Darwiche, 2006a).

We show in this paper that the simplified models can also be used to approximate the partition function if one applies a correction for each deleted edge. We propose two edge-correction schemes, each of which is capable of perfectly correcting the partition function when a single edge has been deleted. The first scheme will have this property only when a particular condition holds in the simplified model, and gives rise to the Bethe free energy approximation when applied to a simplified model that has a tree structure. The second correction scheme does not require such a condition and is shown empirically to lead to more accurate approximations. Both schemes can be applied to the whole spectrum of simplified models and can therefore be used to trade-off the quality of obtained approximations with the complexity of computing them.

## 2 Approximation by Edge Deletion

We first review our edge deletion framework in the context of pairwise Markov random fields (Choi & Darwiche, 2007). For an application to directed models, see (Choi & Darwiche, 2006a).

Let a pairwise Markov random field (MRF) $\mathcal{M}$ have a graph $(\mathcal{E}, \mathcal{V})$ with edges $(i, j) \in \mathcal{E}$ and nodes $i \in \mathcal{V}$, where each node $i$ of the graph is associated with a variable $X_i$ taking on values $x_i$. Edges $(i, j)$ are associated with edge potentials $\psi(x_i, x_j)$ and nodes $i$ with node potentials $\psi(x_i)$. The

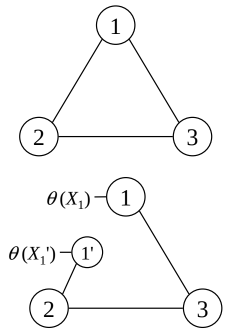


Figure 1: An MRF (top); after edge deletion (bottom).

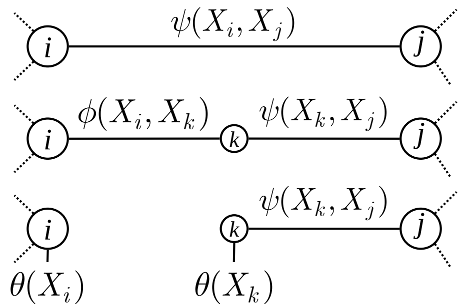


Figure 2: To delete edge $(i, j)$ (top), we introduce auxiliary node $k$ (middle), and delete equivalence edge $(i, k)$, adding edge parameters (bottom).

(strictly positive) distribution $Pr$ induced by $\mathcal{M}$ is defined as follows:

$$Pr(\mathbf{x}) \stackrel{def}{=} \frac{1}{Z} \prod_{(i,j)\in\mathcal{E}} \psi(x_i, x_j) \prod_{i\in\mathcal{V}} \psi(x_i),$$

where $\mathbf{x}$ is an instantiation $x_1, \ldots, x_n$ of network variables, and where $Z$ is the *partition function:*

$$Z \stackrel{def}{=} \sum_{\mathbf{x}} \prod_{(i,j)\in\mathcal{E}} \psi(x_i, x_j) \prod_{i\in\mathcal{V}} \psi(x_i).$$

The basic idea behind our framework is to delete enough edges from the pairwise MRF to render it tractable for exact inference.

**Definition 1** *Let $\mathcal{M}$ be a pairwise MRF. To delete edge $(i, j)$ from $\mathcal{M}$ we remove the edge $(i, j)$ from $\mathcal{M}$ and then introduce the auxiliary potentials $\theta(X_i)$ and $\theta(X_j)$ for variables $X_i$ and $X_j$.*

Figure 1 provides an example of deleting an edge. When deleting multiple edges, we may introduce multiple, yet distinct, potentials $\theta(X_i)$ for the same node $X_i$. We shall refer to auxiliary potentials $\theta(X_i)$ and $\theta(X_j)$ as *edge parameters* and use $\Theta$ to denote the set of all edges parameters. The resulting pairwise MRF will be denoted by $\mathcal{M}'(\Theta)$, its partition function will be denoted by $Z'(\Theta)$ and its distribution will be denoted by $Pr'(.;\Theta)$. When choosing a particular value for edge parameters $\Theta$, we will drop reference to $\Theta$, using only $\mathcal{M}'$, $Z'$ and $Pr'(.)$.

Note that while the distribution $Pr(.)$ and partition function $Z$ of the original pairwise MRF $\mathcal{M}$ may be hard to compute, the distribution $Pr'(.;\Theta)$ and partition function $Z'(\Theta)$ of $\mathcal{M}'(\Theta)$ should be easily computable due to edge deletion. Note also that before we can use $Pr'(.;\Theta)$ and $Z'(\Theta)$ to approximate $Pr(.)$ and $Z$, we must first specify the edge parameters $\Theta$. In fact, it is the values of these parameters which will control the quality of approximations $Pr'(.;\Theta)$ and $Z'(\Theta)$.

Without loss of generality, we will assume that we are only deleting *equivalence edges* $(i, j)$, which connect two variables $X_i$ and $X_j$ with the same domain, and have a potential $\phi(x_i, x_j)$ that denotes an equivalence constraint: $\phi(x_i, x_j) = 1$ if $x_i = x_j$, and $\phi(x_i, x_j) = 0$ otherwise. The deletion of any edge in an MRF can be formulated as the deletion of an equivalence edge.[1]

In previous work, we proposed (and justified) the following conditions on edge parameters $\theta(x_i)$ and $\theta(x_j)$:

$$\theta(x_i) = \alpha \frac{\partial Z'}{\partial \theta(x_j)} \qquad \text{and} \qquad \theta(x_j) = \alpha \frac{\partial Z'}{\partial \theta(x_i)} \tag{1}$$

[1]To delete an MRF edge $(i, j)$ that is not an equivalence edge, we use the technique in Figure 2: we introduce an auxiliary node $k$ between $i$ and $j$; introduce an equivalence constraint on the edge $(i, k)$; copy the original potential of edge $(i, j)$ to $(k, j)$; and delete the equivalence edge $(i, k)$. Note that the original model and the extended one will have the same treewidth in this case, will agree on the distribution over their common variables, and will have the same partition function values.

where $\alpha$ is a normalizing constant. Note that the partial derivatives of Equation 1 can be computed efficiently in traditional inference frameworks (Darwiche, 2003; Park & Darwiche, 2004).

Equation 1 can also be viewed as update equations, suggesting an iterative method that searches for edge parameters, which we called ED-BP (Choi & Darwiche, 2006a). Starting with an initial approximation $\mathcal{M}'_0$ at iteration $t = 0$ (say, with uniform parameters), we can compute edge parameters $\theta_t(x_i)$ and $\theta_t(x_j)$ for an iteration $t > 0$ by performing exact inference in the approximate network $\mathcal{M}'_{t-1}$. We repeat this process until we observe that all parameters converge to a fixed point satisfying Equation 1 (if ever).

Note that Equation 1 does not specify a unique value of edge parameters, due to the constants $\alpha$. That is, each value of these constants will lead to a different set of edge parameters. Yet, independent of which constants we use, the resulting pairwise MRF $\mathcal{M}'$ will have an *invariant* distribution $Pr'(.)$ that satisfies the following properties. First,

$$Pr'(x_i) = Pr'(x_j) = \frac{1}{z_{ij}} \cdot \theta(x_i)\theta(x_j), \tag{2}$$

where $z_{ij} = \sum_{x_i = x_j} \theta(x_i)\theta(x_j)$. Next, if the pairwise MRF $\mathcal{M}'$ has a tree structure, the node marginals of distribution $Pr'(.)$ will correspond precisely to the node marginals obtained by running IBP on the original model $\mathcal{M}$. Finally, if the pairwise MRF $\mathcal{M}'$ has loops (i.e., not a tree), the node marginals of distribution $Pr'$ will correspond to node marginals obtained by running a generalization of IBP on the original model $\mathcal{M}$ (in particular, a class of iterative joingraph propagation algorithms; see Choi & Darwiche, 2006a; Aji & McEliece, 2001; Dechter, Kask, & Mateescu, 2002).

## 3 Approximating the Partition Function by Edge Correction

While the edge parameters specified by Equation 1 are guaranteed to yield an invariant distribution $Pr'(.)$, they are not guaranteed to yield an invariant partition function $Z'$ as this function is sensitive to the choice of constants $\alpha$. While these edge parameters will yield an interesting approximation of node marginals, they do not yield a meaningful approximation of the partition function.

We will show in this section, however, that one can apply an edge-by-edge correction to the partition function $Z'$, leading to a corrected partition function that is invariant to the choice of constants $\alpha$. This seemingly subtle approach leads to two important consequences. First, it results in a new semantics for the Bethe free energy approximation as a corrected partition function. Second, it allows for an improved class of approximations based on improved corrections.

### 3.1 An Edge Correction that Gives Rise to the Bethe Free Energy

We will now propose a correction to the partition function $Z'$, which gives rise to the Bethe free energy and some of its generalizations.

**Proposition 1** *Let $\mathcal{M}'$ be the result of deleting a single equivalence edge $(i, j)$ from a pairwise MRF $\mathcal{M}$. If the parameters of edge $(i, j)$ are given by Equation 1, and if the mutual information between $X_i$ and $X_j$ is zero in $\mathcal{M}'$, then:*

$$Z = Z' \cdot \frac{1}{z_{ij}}, \qquad \textit{where} \qquad z_{ij} = \sum_{x_i = x_j} \theta(x_i)\theta(x_j).$$

That is, if we delete a *single* edge $(i, j)$ and find that $X_i$ and $X_j$ are independent in the resulting model $\mathcal{M}'$, we can correct the partition function $Z'$ by $z_{ij}$ and recover the exact partition function $Z$. Moreover, the result of this correction is invariant to the constants $\alpha$ used in Equation 1.

From now on, we will use $MI(X_i; X_j)$ to denote the mutual information between two variables $X_i$ and $X_j$ in the pairwise MRF $\mathcal{M}'$. Moreover, when $MI(X_i; X_j) = 0$, we will say that the deleted edge $(i, j)$ is a zero-MI edge.

Let us now consider the more realistic situation where we delete multiple edges, say, $\mathcal{E}^\star$, from $\mathcal{M}$ to yield the model $\mathcal{M}'$. One can try to apply the above correction to each of the deleted edges, leading to a corrected partition function $Z' \cdot \frac{1}{z}$, where

$$z = \prod_{(i,j)\in\mathcal{E}^\star} z_{ij} = \prod_{(i,j)\in\mathcal{E}^\star} \sum_{x_i=x_j} \theta(x_i)\theta(x_j). \tag{3}$$

We will refer to this correction as a *zero-MI edge correction*, or just ZERO-EC. This correction is no longer guaranteed to recover the exact partition function $Z$, even if each of the deleted edges is a zero-MI edge. Yet, if the pairwise MRF $\mathcal{M}'$ has a tree structure, applying this correction to the partition function $Z'$ gives rise to the Bethe free energy approximation.

To review, the Bethe free energy $F_\beta$ is an approximation of the true free energy $F$ of a pairwise MRF $\mathcal{M}$, and is exact when $\mathcal{M}$ has a tree structure (Yedidia, Freeman, & Weiss, 2005). Since $F = -\log Z$, we can in principle use $F_\beta$ as an approximation of the partition function $Z$, even when $\mathcal{M}$ does not have a tree structure, i.e., we can use $Z_\beta = \exp\{-F_\beta\}$.

**Theorem 1** *Let $\mathcal{M}'$ be the result of deleting equivalence edges from a pairwise MRF $\mathcal{M}$. If $\mathcal{M}'$ has a tree structure and its edge parameters are as given by Equation 1, we have $Z_\beta = Z' \cdot \frac{1}{z}$.*

Hence, the Bethe approximation of $Z$ is a degenerate case of the ZERO-EC correction. That is, it corresponds to an ZERO-EC correction applied to a pairwise MRF $\mathcal{M}'$ that has a tree structure. Since the ZERO-EC correction is specified purely in quantities available in the model $\mathcal{M}'$, it will be easily computable as long as the model $\mathcal{M}'$ is sparse enough (i.e., it has a treewidth which is manageable under the given computational resources). Hence, this correction can be practically applicable even if $\mathcal{M}'$ does not have a tree structure. In such a case, the correction will lead to an approximation of the partition function that we expect to be superior to the one obtained by the Bethe free energy. We will illustrate this point empirically later in the paper.

### 3.2 Improved Edge Corrections

Proposition 1 gives us a condition that allows us to correct the partition function exactly, but under the assumption that the single edge deleted is zero-MI. The following result allows us, in fact, to correct the partition function when deleting *any* single edge.

**Proposition 2** *Let $\mathcal{M}'$ be the result of deleting a single equivalence edge $(i,j)$ from a pairwise MRF $\mathcal{M}$. If the parameters of edge $(i,j)$ are given by Equation 1, then:*

$$Z = Z' \cdot \frac{y_{ij}}{z_{ij}}, \qquad \textit{where} \qquad y_{ij} = \sum_{x_i=x_j} Pr'(x_i \mid x_j) = \sum_{x_i=x_j} Pr'(x_j \mid x_i).$$

Contrary to the first correction, this one may either increase or decrease the value of $Z'$. Moreover, when the deleted edge $(i,j)$ happens to be zero-MI, Proposition 2 reduces to Proposition 1.

We can also use this proposition as a basis for correcting the partition function when multiple edges are deleted, just as we did in Equation 3. In particular, we now propose using the correction $Z' \cdot \frac{y}{z}$, where $z$ is the same factor given in Equation 3, and

$$y = \prod_{(i,j)\in\mathcal{E}^\star} y_{ij} = \prod_{(i,j)\in\mathcal{E}^\star} \sum_{x_i=x_j} Pr'(x_i \mid x_j) \tag{4}$$

We refer to this correction as a *general edge correction*, or just GENERAL-EC.

## 4 Edge Recovery

Suppose we already have a tree-structured approximation $\mathcal{M}'$ of the original model $\mathcal{M}$, but we are afforded more computational resources. We can then improve the approximation by *recovering* some of the deleted edges. However, which edge's recovery would have the most positive impact on the quality of the approximation?

**Edge Recovery for ZERO-EC.** Since ZERO-EC is exact when a single edge is deleted with $MI(X_i; X_j) = 0$, one may want to recover those edges $(i,j)$ with the highest mutual information $MI(X_i; X_j)$.[2] We will indeed show the promise of this heuristic in Section 6 for ZERO-EC corrections. We will also show that it turns out to be a poor heuristic for GENERAL-EC corrections.

[2]This heuristic was first proposed in (Choi & Darwiche, 2006a).

**Edge Recovery for GENERAL-EC.** Consider the situation when two equivalence edges are deleted, $(i, j)$ and $(s, t)$. Here, we use the approximate correction: $Z' \cdot \frac{y}{z} = Z' \cdot \frac{y_{ij}}{z_{ij}} \frac{y_{st}}{z_{st}}$, where $\frac{y_{ij}}{z_{ij}}$ is the single-edge correction for edge $(i, j)$ and $\frac{y_{st}}{z_{st}}$ is the single-edge correction for edge $(s, t)$.

The question now is: when is this double-edge correction exact? Intuitively, we want to identify a situation where each edge can be corrected, independently of each other. Consider then the case where variables $X_i, X_j$ are independent of variables $X_s, X_t$.

**Proposition 3** *Let $\mathcal{M}'$ be the result of deleting two equivalence edges, $(i, j)$ and $(s, t)$, from a pairwise MRF $\mathcal{M}$. If the edge parameters of $\mathcal{M}'$ are as given by Equation 1, and if $MI(X_i, X_j; X_s, X_t) = 0$ in $\mathcal{M}'$, then:*

$$Z = Z' \cdot \frac{y_{ij}}{z_{ij}} \frac{y_{st}}{z_{st}}.$$

This suggests a new edge recovery heuristic for GENERAL-EC approximations to the partition function. First, we start with a tree-structured network $\mathcal{M}'$, and assign each deleted edge $(i, j)$ a score:

$$\sum_{(s,t) \in \mathcal{E}^\star \setminus (i,j)} MI(X_i, X_j; X_s, X_t).$$

We then prefer to recover the top $k$ with the highest mutual information scores.

## 5 Edge Corrections and Free Energies

When the model $\mathcal{M}'$ is a tree, ZERO-EC yields the influential Bethe free energy approximation (Yedidia et al., 2005). When the model $\mathcal{M}'$ has cycles, it can be shown that ZERO-EC corresponds more generally to joingraph free energy approximations (Aji & McEliece, 2001; Dechter et al., 2002); see (Choi & Darwiche, 2006a) for the connection to iterative joingraph propagation.

GENERAL-EC can also take the form of another free energy approximation. Note that when multiple equivalence edges are deleted, we can compute the partition function $Z'_{ij}$ of a model $\mathcal{M}'_{ij}$ by treating the single edge $(i, j)$ as if it was the only edge deleted: $Z'_{ij} = Z' \cdot \frac{y_{ij}}{z_{ij}}$. Therefore, we have that:

$$Z' \cdot \frac{y}{z} = Z' \cdot \prod_{(i,j) \in \mathcal{E}^\star} \frac{y_{ij}}{z_{ij}} = Z' \cdot \prod_{(i,j) \in \mathcal{E}^\star} \frac{Z'_{ij}}{Z'}.$$

This leads to a free energy of the form $-\log(Z' \cdot \frac{y}{z}) = (n - 1) \log Z' - \sum_{(i,j) \in \mathcal{E}^\star} \log Z'_{ij}$, where $n$ is the number of equivalence edges $(i, j)$ deleted. The structure of this (dual) free energy can be visualized via an EP graph (Welling, Minka, & Teh, 2005), which can be used to design EP free energy approximations (Minka, 2001a, 2001b), consisting of $\mathcal{M}'$ as the base approximation with models $\mathcal{M}'_{ij}$ as outer regions. Whereas we fixed, somewhat arbitrarily, our edge parameters to satisfy Equation 1, we could in principle seek edge parameters optimizing the above free energy directly (see the Appendix), lifting edge deletion approximations towards EP and GBP free energy approximations with higher-order structure. In the other direction, this perspective could provide new semantics for certain EP and GBP free energies. Moreover, the edge recovery heuristic from the previous section can serve as a heuristic for identifying improved EP and GBP free energies. This is a perspective that is currently being investigated.

Although much of the progress in IBP and its generalizations have relied on the free energy perspective, yielding successful algorithms for approximating marginals, relatively little attention has been paid to direct approximations of the partition function. A notable exception is tree-reweighted belief propagation (Wainwright, Jaakkola, & Willsky, 2005), which provides upper bounds on the log partition function that can be thought of as a convexified form of the Bethe free energy. Mean field methods and its generalizations are another well-known class of approximations that provide lower bounds on the partition function (e.g., Saul & Jordan, 1995; Jaakkola, 2000). These, and other variational approximations (Jordan, Ghahramani, Jaakkola, & Saul, 1999; Choi & Darwiche, 2006b), provide a complementary approach to approximate inference as exact inference in simplified models, but one based on solving a nonlinear optimization problem. In Section 6, we highlight an initial comparison of edge-corrected partition function approximations against a generalization of the mean field approximations given by (Wiegerinck, 2000; Geiger, Meek, & Wexler, 2006).

## 6 Experimental Results

Our goal here is to highlight the effects that different edge recovery heuristics have on ZERO-EC and GENERAL-EC approximations. For each plot, we generate over 50 problem instances, omitting those instances where ED-BP fails to converge in 100 iterations. Starting from a random spanning tree, we rank each deleted edge, and recover edges $k$ at a time until all edges are recovered. At each point, we evaluate the quality of the approximation by the average relative error $|\widehat{Z} - Z|/Z$, where $\widehat{Z}$ denotes the appropriate approximation. Note that when no edges are recovered (from the tree approximation), ZERO-EC corresponds to the Bethe approximation.

In Figure 3, we generated random $6 \times 6$ grid networks, where pairwise couplings were given random parameters (left), or random parameters in $[0.0, 0.1)$ or $(0.9, 1.0]$ (right). We recovered edges randomly here, and compared ZERO-EC (Z-EC) and GENERAL-EC (G-EC) with variational lower bounds on $Z$ based on a mean field generalization (Wiegerinck, 2000; Geiger et al., 2006). We gave the variational approximation the same network structure as ZERO-EC and GENERAL-EC (eliminating the auxiliary variables). We find that the variational bounds are loose here, compared to the approximations given by ZERO-EC and GENERAL-EC.

In Figure 4, we compare ZERO-EC approximations with GENERAL-EC approximations. In each of the four plots, we use random edge recovery as a baseline. In the first plot from the left, we evaluated the (static) edge recovery heuristic used in (Choi & Darwiche, 2006a), where we recovered edges first with the highest mutual information (MI+) score in the tree approximation. We see that this heuristic has a positive effective on the quality of ZERO-EC approximations, relative to random recovery. In the second plot, we reversed the decisions, choosing to recover first those with the lowest score (MI–), and see the opposite effect. In both plots, the heuristic appears to have little effect on the GENERAL-EC approximations. In the last two plots of Figure 4, we evaluate the heuristics proposed in Section 4 in a similar way (MI2+ and MI2–). We see now this new heuristic has a positive effect on GENERAL-EC, and ZERO-EC as well. Note also that GENERAL-EC dominates the respective ZERO-EC approximation, on average.

As all of the previous results can be formulated for Bayesian networks in a straightforward manner, we also run experiments here on Bayesian networks, for which a variety of benchmark networks are publicly available.[3] Here each problem instance is an instantiation of evidence, randomly set on all leaves of the network.

In Figure 5, we see similar results in the `win95pts` and `water` networks, as we saw in our random grid networks. In Figure 6, the `alarm` network appears to show that ZERO-EC approximations may not be sensitive in some networks to the edge recovery heuristic introduced for GENERAL-EC.

## 7 Conclusion

We proposed an approach for approximating the partition function which is based on two steps: (1) computing the partition function of a simplified model which is obtained by deleting model edges, and (2) rectifying the result by applying an edge-by-edge correction. The approach leads to an intuitive framework in which one can trade-off the quality of an approximation with the complexity of computing it through a simple process of edge recovery. We provided two concrete instantiations of the proposed framework by proposing two edge correction schemes with corresponding heuristics for edge recovery. The first of these instantiations captures the well known Bethe free energy approximation as a degenerate case. The second instantiation has been shown to lead to more accurate approximations and appears to hold promise for capturing a more sophisticated class of free energy approximations.

[3] The networks used here are available at http://www.cs.huji.ac.il/labs/compbio/Repository.

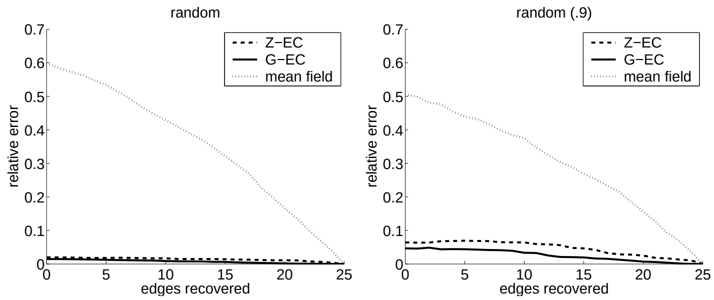


Figure 3: Edge correction in random $6 \times 6$ grid networks.

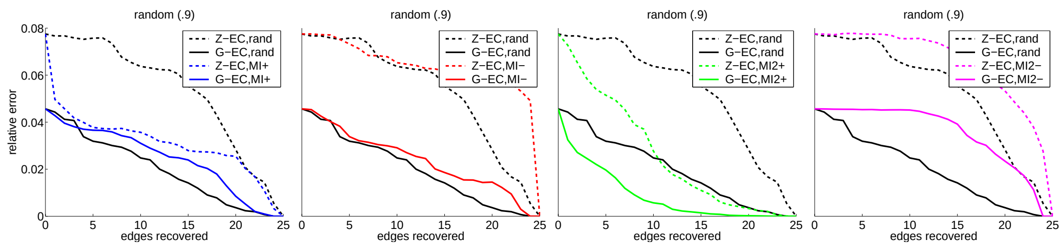


Figure 4: Edge correction in random $6 \times 6$ grid networks.

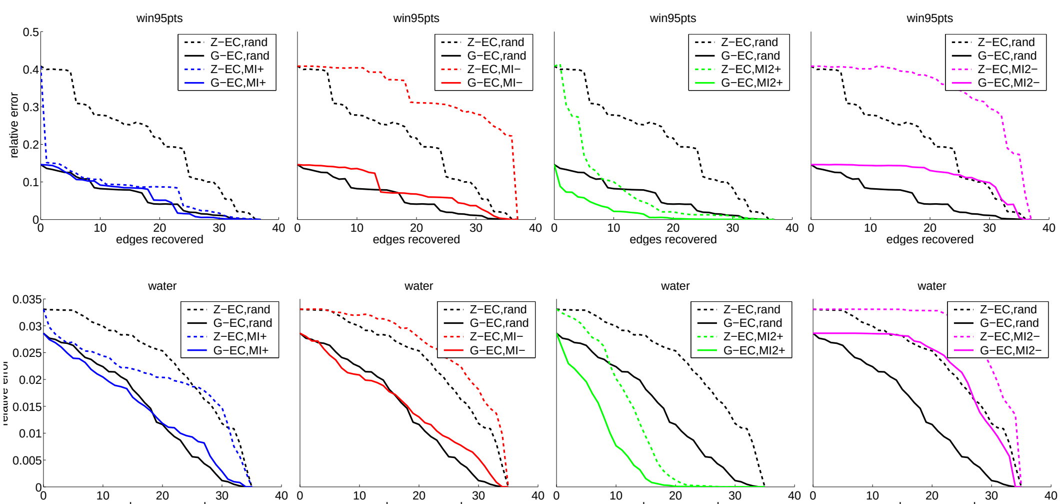


Figure 5: Edge correction in `win95pts` and `water` networks.

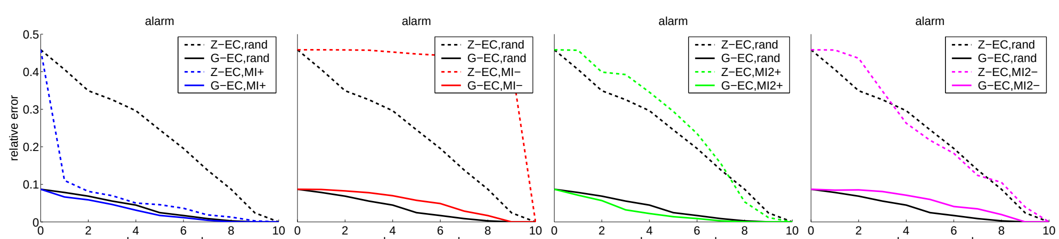


Figure 6: Edge correction in the `alarm` network.

## A Stationary point conditions for general edge correction

We consider here stationary points of

$$f = -\log \big[ Z' \cdot \prod_{(i,j)\in\mathcal{E}^\star} \frac{Z'_{ij}}{Z'} \big] = -\log Z' - \sum_{(i,j)\in\mathcal{E}^\star} \big[ \log Z'_{ij} - \log Z' \big].$$

Consider the partial derivatives of $f$ with respect to $\theta(x_i)$:

$$\begin{aligned}
\frac{\partial f}{\partial \theta(x_i)} &= \frac{\partial}{\partial \theta(x_i)} \big\{ -\log Z' - \big[ \log Z'_{ij} - \log Z' \big] - \sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} \big[ \log Z'_{st} - \log Z' \big] \big\} \\
&= \frac{\partial}{\partial \theta(x_i)} \big\{ -\log Z'_{ij} - \sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} \big[ \log Z'_{st} - \log Z' \big] \big\} \\
&= 0 - \sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} \Big[ \frac{1}{Z'_{st}} \frac{\partial Z'_{st}}{\partial \theta(x_i)} - \frac{1}{Z'} \frac{\partial Z'}{\partial \theta(x_i)} \Big]
\end{aligned}$$

Setting these partial derivatives to zero, and rearranging we have:

$$\sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} \frac{1}{Z'} \frac{\partial Z'}{\partial \theta(x_i)} = \sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} \frac{1}{Z'_{st}} \frac{\partial Z'_{st}}{\partial \theta(x_i)}.$$

Multiplying by $\theta(x_i)$ on each side, and simplifying, we have stationary point conditions:

$$\sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} Pr'(x_i) = \sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} Pr'_{st}(x_i),$$

or simply

$$\sum_{(s,t)\in\mathcal{E}^\star\setminus(i,j)} (n-1) Pr'(x_i) = Pr'_{st}(x_i), \tag{5}$$

where $n$ is the number of edges deleted. Analagously for edge parameters $\theta(x_j)$.

One condition where these conditions are satisfied is when all networks $\mathcal{M}'$ and $\mathcal{M}'_{ij}$ agree on marginals for variable $X_i$ and $X_j$. However, Condition 5 is weaker, and asks only that the sum of the marginals for $x_i$ be consistent.

### References


Aji, S. M., & McEliece, R. J. (2001). The generalized distributive law and free energy minimization. In *Proceedings of the 39th Allerton Conference on Communication, Control and Computing*, pp. 672–681.

Choi, A., & Darwiche, A. (2006a). An edge deletion semantics for belief propagation and its practical impact on approximation quality. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pp. 1107–1114.

Choi, A., & Darwiche, A. (2006b). A variational approach for approximating Bayesian networks by edge deletion. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 80–89.

Choi, A., & Darwiche, A. (2007). An edge deletion semantics for belief propagation.. Under review.

Darwiche, A. (2003). A differential approach to inference in bayesian networks. *Journal of the ACM*, *50*(3), 280–305.

Dechter, R., Kask, K., & Mateescu, R. (2002). Iterative join-graph propagation. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 128–136.

Geiger, D., Meek, C., & Wexler, Y. (2006). A variational inference procedure allowing internal structure for overlapping clusters and deterministic constraints. *J. Artif. Intell. Res. (JAIR)*, *27*, 1–23.

Jaakkola, T. (2000). Tutorial on variational approximation methods. In Opper, M., & Saad, D. (Eds.), *Advanced mean field methods: theory and practice.*, pp. 129–159, Cambridge, Massachusetts.

Jordan, M. I., Ghahramani, Z., Jaakkola, T., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*(2), 183–233.

Minka, T. P. (2001a). The EP energy function and minimization schemes. Tech. rep., Carnegie Mellon University.

Minka, T. P. (2001b). *A family of algorithms for approximate Bayesian inference*. Ph.D. thesis, MIT.

Park, J., & Darwiche, A. (2004). A differential semantics for jointree algorithms. *Artificial Intelligence*, *156*, 197–216.

Saul, L. K., & Jordan, M. I. (1995). Exploiting tractable substructures in intractable networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 486–492.

Wainwright, M. J., Jaakkola, T., & Willsky, A. S. (2005). Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, *51*(11), 3697–3717.

Welling, M., Minka, T. P., & Teh, Y. W. (2005). Structured region graphs: morphing EP into GBP. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*.

Wiegerinck, W. (2000). Variational approximations between mean field theory and the junction tree algorithm. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 626–633.

Yedidia, J., Freeman, W., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, *51*(7), 2282–2312.