Fixing the Bethe Approximation: How Structural Modifications in a Graph Improve Belief Propagation – Supplementary Material

Harald Leisenberger[1] **Franz Pernkopf**[1] **Christian Knoll**[1]

[1]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

APPENDIX

APPENDIX A – PROOFS

PROOF OF LEMMA 1:.

Inserting $q_i = 0.5$ and $q_j = 0.5$ in (16) yields

$$\begin{aligned}
\xi^*_{ij}(0.5, 0.5) &= \frac{1}{2\alpha_{ij}}\Big((1+\alpha_{ij}) - \sqrt{(1+\alpha_{ij})^2 - \alpha_{ij}(1+\alpha_{ij})}\ \Big) \\
&= \frac{1}{2\alpha_{ij}}\Big((1+\alpha_{ij}) - \sqrt{(1+\alpha_{ij})}\ \Big) \\
&\overset{(15)}{=} \frac{1}{2\big(e^{4J_{ij}}-1\big)}\big(e^{4J_{ij}} - \sqrt{e^{4J_{ij}}}\big) \\
&= \frac{1}{2}\Big(\frac{e^{2J_{ij}}-1}{e^{2J_{ij}}-e^{-2J_{ij}}}\Big) \\
&= \frac{1}{4}\big(\tanh(J_{ij})+1\big) \\
&= \frac{\sigma(2\,J_{ij})}{2}.
\end{aligned}$$

□

PROOF OF LEMMA 2:.

Note first that one can interchange the role of q_i and q_j in the definition (16) of ξ^*_{ij} due to symmetry. It is therefore sufficient to consider only the left hand side equalities in (22) and (23). Let first $q_i \to 0$ and $q_j \to k$. From (9) we know that ξ^*_{ij} is bounded from above by $\min(q_i, q_j)$. As also $\xi^*_{ij} > 0$, the first equality in (22) follows by continuity. Let now $q_i \to 1$ and $q_j \to k$. Then the limit of both the lower bound and the upper bound in (9) equals k. Consequently, ξ^*_{ij} must tend to k as well, which yields the first equality in (23). □

PROOF OF LEMMA 3:.

(a) We apply Lemma 1 of Dragomir et al. [2000] to the special case of binary random variables. For the upper bound, we substitute the matching combinations of singleton and pairwise probabilities from table (1) into the right-hand formula from (23):

$$I_B^{(i,j)}(q_i, q_j) \;\leq\; \frac{(\xi^*_{ij})^2}{q_i q_j} + \frac{(q_i - \xi^*_{ij})^2}{q_i(1-q_j)} + \frac{(q_j - \xi^*_{ij})^2}{(1-q_i)q_j} + \frac{(1+\xi^*_{ij} - q_i - q_j)^2}{(1-q_i)(1-q_j)} - 1,$$

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

and after a few simple algebraic manipulations we directly arrive at the desired result. Analogously, we derive the lower bound by inserting the corresponding probabilities into the expression to the left of $I_B^{(i,j)}$ in (23):

$$\begin{aligned} I_B^{(i,j)}(q_i,q_j) \;\geq\; \frac{1}{2}\big(\,|\,\xi_{ij}^* - q_iq_j\,|\,+|\,(1+\xi_{ij}^*-q_i-q_j)-(1-q_i)(1-q_j)\,|\\ +|\,(q_i-\xi_{ij}^*)-q_i(1-q_j)\,|+|\,(q_j-\xi_{ij}^*)-(1-q_i)q_j\,|\,\big)^2 \end{aligned}$$

Depending on whether we have an attractive or an repulsive edge, we either make us of result (a) or (b) in Lemma 2 and get rid of the absolute value symbols, after which – in both cases – the last expression simplifies to

$$\frac{1}{2}\big(4(\xi_{ij}^* - q_iq_j)\big)^2 = 8(\xi_{ij}^* - q_iq_j)^2.$$

(b) According to 20, the boundary of the sliced Bethe Box $\mathbb{B}^{(i,j)}$ is the union of four line segments

$$\begin{aligned} \partial\mathbb{B}^{(i,j)} =&\{(q_i,q_j)\in\mathbb{R}^2\,|\,q_i=0, 0\leq q_j\leq 1\}\cup\\ &\{(q_i,q_j)\in\mathbb{R}^2\,|\,q_i=1, 0\leq q_j\leq 1\}\cup\\ &\{(q_i,q_j)\in\mathbb{R}^2\,|\,0\leq q_i\leq 1, q_j=0\}\cup\\ &\{(q_i,q_j)\in\mathbb{R}^2\,|\,0\leq q_i\leq 1, q_j=1\}. \end{aligned}$$

Without loss of generality, we focus on the case where $q_i\to 1$ and $q_j\to k$ for some $k\in[0,1]$. The remaining cases can be treated similarly. We further note that it is sufficient to prove the equality

$$\lim_{\substack{q_i\to 1\\ q_j\to k}} \tilde{P}_{ij}(x_i,x_j) = \lim_{\substack{q_i\to 1\\ q_j\to k}} \tilde{P}_i(x_i)\tilde{P}_j(x_j) \tag{1}$$

for all possible realizations $x_i, x_j\in\{+1,-1\}$ of X_i, X_j, as this implies statistical independence of X_i and X_j at the boundary and consequently their mutual information must equal zero in the limit. For checking equality (1), we utilize Lemma 2. E.g., for $x_i=+1, x_j=-1$ we get

$$\begin{aligned} \lim_{\substack{q_i\to 1\\ q_j\to k}} \tilde{P}_{ij}(+1,-1) &\overset{(1)}{=} \lim_{\substack{q_i\to 1\\ q_j\to k}} (q_i - \xi_{ij}^*(q_i,q_j))\\ &\overset{(23)}{=} 1-k =\\ &= \lim_{\substack{q_i\to 1\\ q_j\to k}} q_i(1-q_j)\\ &\overset{(1)}{=} \lim_{\substack{q_i\to 1\\ q_j\to k}} \tilde{P}_i(+1)\tilde{P}_j(-1). \end{aligned}$$

Analogous calculations for the three other possible combinations of x_i, x_j validate the statement.

□

PROOF OF LEMMA 4:.

(a) We utilize Lemma 3 of Welling and Teh [2001]. Then the first-order derivative of $\mathcal{F}_B^{\setminus(i,j)}$ on $\mathbb{B}^{\setminus(i,j)}$ is

$$\frac{\partial}{\partial q_i}\mathcal{F}_B^{\setminus(i,j)} = -2\,\theta_i + 2\sum_{k\in\mathcal{N}(i)\setminus j} J_{ij} + \log\Big(\frac{(1-q_i)^{d_i-2}}{q_i^{d_i-2}}\prod_{k\in\mathcal{N}(i)\setminus j}\frac{q_i-\xi_{ij}^*}{1+\xi_{ij}^*-q_i-q_j}\Big). \tag{2}$$

Consequently, we obtain

$$\begin{aligned} \frac{\partial}{\partial q_i}\Delta\mathcal{F}_B^{(i,j)} &= \frac{\partial}{\partial q_i}(\mathcal{F}_B - \mathcal{F}_B^{\setminus(i,j)})\\ &= \frac{\partial}{\partial q_i}\mathcal{F}_B - \frac{\partial}{\partial q_i}\mathcal{F}_B^{\setminus(i,j)}\\ &= 2J_{ij} + \log\Big(\frac{(1-q_i)(q_i-\xi_{ij}^*)}{q_i(1+\xi_{ij}^*-q_i-q_j)}\Big). \end{aligned} \tag{3}$$

(b) We can apply Theorem 1 from Weller and Jebara [2013]. Note, however, that we must also take the second derivatives of the node entropies $\mathcal{S}_i$ and $\mathcal{S}_j$ into account which are computed as

$$\frac{\partial^2}{\partial q_i^2}\,\mathcal{S}_i = -\frac{1}{q_i(1-q_i)}, \tag{4}$$

$$\frac{\partial^2}{\partial q_j^2}\,\mathcal{S}_j = -\frac{1}{q_j(1-q_j)}, \tag{5}$$

and are zero for the cross derivatives. By definition, we have $\Delta\mathcal{F}_B^{(i,j)} = \Delta\mathcal{U}_B^{(i,j)} + I_B^{(i,j)} = \Delta\mathcal{U}_B^{(i,j)} + \mathcal{S}_i + \mathcal{S}_j - \mathcal{S}_{ij}$. Note that f_{ij} from Theorem 1 in Weller and Jebara [2013]corresponds precisely to $\Delta\mathcal{U}_B^{(i,j)} - \mathcal{S}_{ij}$. If we put these observations together, the statement follows immediately.

□

PROOF OF LEMMA 5:.

The proof consists of two parts: in the first part, we show that $(0.5, 0.5)$ is the only stationary point of $\Delta\mathcal{F}_B^{(i,j)}$ on the sliced Bethe box. In the second part, we show that the Hessian matrix $\nabla^2\mathcal{F}_B$ evaluated in $(0.5, 0.5)$ is indefinite.

Part 1: Setting the gradient $\nabla\,\Delta\mathcal{F}_B^{(i,j)}$ with its components given by (3) to zero, leads to the following nonlinear equation system:

$$2J_{ij} + \log\Big(\frac{q_i - \xi_{ij}^* - q_i^2 + q_i\xi_{ij}^*}{q_i + q_i\xi_{ij}^* - q_i^2 - q_iq_j}\Big) = 0$$

$$2J_{ij} + \log\Big(\frac{q_j - \xi_{ij}^* - q_j^2 + q_j\xi_{ij}^*}{q_j + q_j\xi_{ij}^* - q_j^2 - q_iq_j}\Big) = 0$$

which is equivalent to

$$e^{2J_{ij}}(q_i - \xi_{ij}^* - q_i^2 + q_i\xi_{ij}^*) = q_i + q_i\xi_{ij}^* - q_i^2 - q_iq_j \tag{6}$$

$$e^{2J_{ij}}(q_j - \xi_{ij}^* - q_j^2 + q_j\xi_{ij}^*) = q_j + q_j\xi_{ij}^* - q_j^2 - q_iq_j. \tag{7}$$

By subtracting (7) from (6), we can reduce the above system to one equation

$$(e^{2J_{ij}} - 1)(q_j + q_i)(q_j - q_i) + (-e^{2J_{ij}}\xi_{ij}^* - e^{2J_{ij}} + \xi_{ij}^* + 1)(q_j - q_i) = 0. \tag{8}$$

Note that (8) might possess additional solutions, that do not solve the original system (6) + (7); each solution of (8), however, does also solve (6) + (7). Obviously all feasible pairs (q_i, q_j) such that $q_i = q_j$ solve equation (8) and thus also (6) + (7). Now assume that there exists a feasible solution to (8) with unequal components. If we divide (8) by $(q_j - q_i)$ and simplify, we end up with the contradictory result

$$\underbrace{(1 + \xi_{ij}^* - q_i - q_j)}_{=\tilde{P}_{ij}(X_i=-1,X_j=-1)>0}\ \underbrace{(1 - e^{2J_{ij}})}_{\neq 0} = 0. \tag{9}$$

Consequently, the only candidates for stationary points of $\Delta\mathcal{F}_B^{(i,j)}$ on $\mathbb{B}^{(i,j)}$ are those with equal components. This allows us to substitute q_i for q_j in either of the two equations (6) + (7) and directly solve for q_i. In other words, we must solve

$$e^{2J_{ij}}(q_i - \xi_{ij}^* - q_i^2 + q_i\xi_{ij}^*) - (q_i + q_i\xi_{ij}^* - 2q_i^2) = 0,$$

or equivalently,

$$\xi_{ij}^* = \frac{(1 - e^{2J_{ij}})q_i + (e^{2J_{ij}} - 2)q_i^2}{-e^{2J_{ij}} - (1 - e^{2J_{ij}})q_i},$$

where we can replace ξ_{ij}^* by formula (16) (with $q_j = q_i$):

$$(1 + 2\alpha_{ij}q_i) - \sqrt{1 + 4\alpha_{ij}q_i(1 - q_i)} = 2\alpha_{ij}\underbrace{\frac{(1 - e^{2J_{ij}})q_i + (e^{2J_{ij}} - 2)q_i^2}{-e^{2J_{ij}} - (1 - e^{2J_{ij}})q_i}}_{:=A}$$

After isolating the radical and squaring both sides of the equation we get

$$(1+\alpha_{ij})q_i^2 - (1+2\alpha_{ij}q_i)A + \alpha_{ij}A^2 = 0,$$

which can be further simplified to

$$\frac{e^{2J_{ij}}(e^{2J_{ij}}-1)(2q_i-1)(q_i-1)^2}{(q_i+e^{2J_{ij}}-q_ie^{2J_{ij}})^2} = 0.$$

Finally, we multiply both sides by $\frac{(q_i+e^{2J_{ij}}-q_ie^{2J_{ij}})^2}{\underbrace{e^{2J_{ij}}}_{\neq 0}\underbrace{(e^{2J_{ij}}-1)}_{\neq 0}\underbrace{(q_i-1)^2}_{\neq 0}}$ and end up with

$$2q_i - 1 = 0 \quad \Rightarrow \quad q_i = 0.5.$$

We conclude that $(\bar{q_i}, \bar{q_j}) = (0.5, 0.5)$ is the only stationary point of $\Delta\mathcal{F}_B^{(i,j)}$.

Part 2: To compute the Hessian matrix $\nabla^2\,\Delta\mathcal{F}_B^{(i,j)}$, we use the second partial derivatives in (27) - (29). Lemma 1 provides us with a simple expression of ξ_{ij}^* evaluated in $(0.5, 0.5)$. Furthermore, we make use of the relations

$$\sigma(2x)(1-\sigma(2x)) \;=\; \frac{1}{2}\sigma'(2x) \;=\; \frac{1}{4\cosh^2(x)} \tag{10}$$

and

$$(1-2\sigma(2x)) \;=\; -\tanh(x). \tag{11}$$

Observe that

$$\begin{aligned} & q_iq_j(1-q_i)(1-q_j) - (\xi_{ij}^* - q_iq_j)^2 \\ =\; & \frac{1}{16} - \Big(\frac{1}{2}\sigma(2\,J_{ij}) - \frac{1}{4}\Big)^2 \\ =\; & \frac{1}{4}\sigma(2\,J_{ij})\big(1-\sigma(2\,J_{ij})\big) \overset{(10)}{=} \frac{1}{16\cosh^2(J_{ij})}. \end{aligned} \tag{12}$$

Then the Hessian evaluated in $(0.5, 0.5)$ is

$$\begin{aligned} & \nabla^2\,\Delta\mathcal{F}_B^{(i,j)}(0.5,0.5) \\ =\; & \begin{pmatrix} \frac{1}{\sigma(2\,J_{ij})\left(1-\sigma(2\,J_{ij})\right)} - 4 & \frac{1-2\sigma(2\,J_{ij})}{\sigma(2\,J_{ij})\left(1-\sigma(2\,J_{ij})\right)} \\ \frac{1-2\sigma(2\,J_{ij})}{\sigma(2\,J_{ij})\left(1-\sigma(2\,J_{ij})\right)} & \frac{1}{\sigma(2\,J_{ij})\left(1-\sigma(2\,J_{ij})\right)} - 4 \end{pmatrix} \\ \overset{(11),(12)}{=}\; & 4\begin{pmatrix} \cosh^2(J_{ij}) - 1 & -\cosh^2(J_{ij})\tanh(J_{ij}) \\ -\cosh^2(J_{ij})\tanh(J_{ij}) & \cosh^2(J_{ij}) - 1 \end{pmatrix} \\ =\; & 4\sinh(J_{ij})\begin{pmatrix} \sinh(J_{ij}) & -\cosh(J_{ij}) \\ -\cosh(J_{ij}) & \sinh(J_{ij}) \end{pmatrix}. \end{aligned}$$

Finally, we compute the leading principal minors of the Hessian, i.e., the determinants of all upper left square submatrices. The first leading principal minor is

$$|4\sinh^2(J_{ij})| \;>\; 0. \tag{13}$$

The second leading principal minor – and thus the determinant of the entire Hessian matrix – is

$$\begin{aligned} & 16\sinh^2(J_{ij})\underbrace{\big(\sinh^2(J_{ij}) - \cosh^2(J_{ij})\big)}_{=-1} \\ =\; & -16\sinh^2(J_{ij}) \;<\; 0. \end{aligned} \tag{14}$$

Since the leading principal minors alternate in sign, starting with a positive number, it follows that the Hessian evaluated in $(0.5, 0.5)$ is indefinite. □

PROOF OF LEMMA 6:.

We only prove the statement for attractive edges as the statement for repulsive edges can be proven analogously. Moreover, we only consider the scenario that both q_i, q_j are > 0.5 (due to symmetry, the reverse scenario can be treated analogously). Let $\mathbb{B}^{(i,j)}_{>0.5} \subseteq \mathbb{B}^{(i,j)}$ be the orthant of $\mathbb{B}^{(i,j)}$ that consists of all points (q_i, q_j) with $0.5 < q_i, q_j < 1$. We know from Lemma 5 that $\Delta\mathcal{F}_B^{(i,j)}$ has no stationary point in $\mathbb{B}^{(i,j)}_{>0.5}$ and is thus bounded from above and below by the limit values of $\Delta\mathcal{F}_B^{(i,j)}$ at the boundary of $\mathbb{B}^{(i,j)}_{>0.5}$. If we can prove that all limit values of $\Delta\mathcal{F}_B^{(i,j)}$ at the boundary of $\mathbb{B}^{(i,j)}_{>0.5}$ are at most zero, then the statement follows immediately (as $\Delta\mathcal{F}_B^{(i,j)}$ is continuous in the interior of $\mathbb{B}^{(i,j)}$).

The boundary of $\mathbb{B}^{(i,j)}_{>0.5}$ consists of four line segments. Let us first consider the line segments that connect $(0.5, 1)$ to $(1, 1)$ and $(1, 0.5)$ to $(1, 1)$. In the proof of Theorem 1 it is shown[1] that

$$\lim_{\substack{q_i \to 0.5 \\ q_j \to 1}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) = 0,$$
$$\lim_{\substack{q_i \to 1 \\ q_j \to 0.5}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) = 0,$$
$$\lim_{\substack{q_i \to 1 \\ q_j \to 1}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) = -J_{ij},$$

and that $\Delta\mathcal{F}_B^{(i,j)}(q_i, q_j)$ is monotonically decreasing from $(0.5, 1)$ to $(1, 1)$ and from $(1, 0.5)$ to $(1, 1)$. Consequently the statement for these two line segments is correct.

Now consider the two remaining line segments, i.e., that connect $(0.5, 0.5)$ to $(0.5, 1)$ and $(0.5, 0.5)$ to $(1, 0.5)$. Due to symmetry, we can again focus on the first case. Let us explicitly compute the value of $\Delta\mathcal{F}_B^{(i,j)}(q_i, q_j)$ in $(0.5, 0.5)$. For that purpose, we must first evaluate the pairwise entropy $\mathcal{S}_{ij}$ in $(0.5, 0.5)$, where we utilize Lemma 1:

$$\begin{aligned}\mathcal{S}_{ij}(0.5, 0.5) =& -\frac{\sigma(2\,J_{ij})}{2}\log(\frac{\sigma(2\,J_{ij})}{2}) + (\frac{\sigma(2\,J_{ij})}{2})\log(\frac{\sigma(2\,J_{ij})}{2}) - 2(0.5 - \frac{\sigma(2\,J_{ij})}{2})\log(0.5 - \frac{\sigma(2\,J_{ij})}{2}) \\ =& -2(0.5 - \frac{\sigma(2\,J_{ij})}{2})\log(0.5 - \frac{\sigma(2\,J_{ij})}{2}).\end{aligned}$$

Then we obtain

$$\begin{aligned}\Delta\mathcal{F}_B^{(i,j)}(0.5, 0.5) &= (-1 - 2(2\xi^*_{ij}(0.5, 0.5) - 0.5 - 0.5))J_{ij} + \overbrace{\mathcal{S}_i(0.5)}^{=\log(2)} + \overbrace{\mathcal{S}_j(0.5)}^{=\log(2)} - \mathcal{S}_{ij}(0.5, 0.5) \\ &= -J_{ij} + 2J_{ij}(1 - \sigma(2J_{ij})) + \log(2) + \sigma(2J_{ij})\log(\frac{\sigma(2J_{ij})}{1 - \sigma(2J_{ij})}) + \log(1 - \sigma(2J_{ij})) \\ &= J_{ij} + \log(2) + \log(1 - \sigma(2J_{ij})) \\ &= \log(2e^{J_{ij}}(1 - \sigma(2J_{ij}))) \\ &= \log(\mathrm{sech}(J_{ij})) < 0.\end{aligned}$$

If we can finally show that $\Delta\mathcal{F}_B^{(i,j)}$ is monotonically increasing from $(0.5, 0.5)$ to $(0.5, 1)$, then the statement of the Lemma follows. By an analogous calculation as in the proof of Lemma 5, we can conclude that the (one-dimensional) function $\Delta\mathcal{F}_B^{(i,j)}(0.5, q_j)$ has no stationary point for $q_j \in (0.5, 1)$ and must therefore be monotonically increasing (as $\Delta\mathcal{F}_B^{(i,j)}(0.5, 0.5) < 0$ and $\lim_{\substack{q_i \to 0.5 \\ q_j \to 1}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) = 0$). □

[1] Note that these results follow independently from the current statement and can therefore be utilized, although Theorem 1 appears chronologically later in this work than Lemma 6.

PROOF OF THEOREM 1:.

By Lemma 5, the Bethe energy difference function does not have any local optima and hence both its greatest lower bound and least upper bound must be located at the boundary of the sliced Bethe box. As in the proof of Lemma 3 (b), we denote this boundary by $\partial\mathbb{B}^{(i,j)}$. Let further $\overline{\mathbb{B}^{(i,j)}} = \mathbb{B}^{(i,j)} \dot\cup \partial\mathbb{B}^{(i,j)}$ be the closure of $\mathbb{B}^{(i,j)}$ and

$$\overline{\Delta\mathcal{F}_B^{(i,j)}}(q_i, q_j) \coloneqq \lim_{(q_i,q_j)\to\overline{\mathbb{B}^{(i,j)}}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) \tag{15}$$

be the analytic continuation of $\Delta\mathcal{F}_B^{(i,j)}$ to $\overline{\mathbb{B}^{(i,j)}}$. Without loss of generality, we assume (i,j) to be an attractive edge (the calculations for a repulsive edge can be done analogously). We utilize Lemma 2 and Lemma 3 (b) to compute the limit values of $\Delta\mathcal{F}_B^{(i,j)}(q_i, q_j)$ at the four corner points of $\overline{\mathbb{B}^{(i,j)}}$:

$$\begin{aligned}
&\lim_{\substack{q_i\to 0\\ q_j\to 0}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) \\
&= \lim_{\substack{q_i\to 0\\ q_j\to 0}} \Delta\mathcal{U}_B^{(i,j)}(q_i, q_j) + \overbrace{\lim_{\substack{q_i\to 0\\ q_j\to 0}} I_B^{(i,j)}(q_i, q_j)}^{=0} \\
&= \lim_{\substack{q_i\to 0\\ q_j\to 0}} -\left(1 + 2\left(2\,\xi_{ij} - q_i - q_j\right)\right) J_{ij}) \\
&= -J_{ij} - \lim_{\substack{q_i\to 0\\ q_j\to 0}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 0} - \overbrace{q_i}^{\to 0} - \overbrace{q_j}^{\to 0})) \\
&= -J_{ij},
\end{aligned}$$

$$\begin{aligned}
&\lim_{\substack{q_i\to 0\\ q_j\to 1}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) \\
&= -J_{ij} - \lim_{\substack{q_i\to 0\\ q_j\to 1}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 0} - \overbrace{q_i}^{\to 0} - \overbrace{q_j}^{\to 1})) \\
&= J_{ij},
\end{aligned}$$

$$\begin{aligned}
&\lim_{\substack{q_i\to 1\\ q_j\to 0}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) \\
&= -J_{ij} - \lim_{\substack{q_i\to 1\\ q_j\to 0}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 0} - \overbrace{q_i}^{\to 1} - \overbrace{q_j}^{\to 0})) \\
&= J_{ij},
\end{aligned}$$

$$\begin{aligned}
&\lim_{\substack{q_i\to 1\\ q_j\to 1}} \Delta\mathcal{F}_B^{(i,j)}(q_i, q_j) \\
&= -J_{ij} - \lim_{\substack{q_i\to 1\\ q_j\to 1}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 2} - \overbrace{q_i}^{\to 1} - \overbrace{q_j}^{\to 1})) \\
&= -J_{ij}.
\end{aligned}$$

For the remainder of the proof, we observe that $\overline{\Delta\mathcal{F}_B^{(i,j)}}$ exhibits monotonic behavior between the corner points of $\overline{\mathbb{B}^{(i,j)}}$.

On the one hand, it is monotonically increasing over the two line segments that connect $(0,0)$ to $(0,1)$ and $(0,0)$ to $(1,0)$:

$$\begin{aligned}\overline{\Delta\mathcal{F}_B^{(i,j)}}(0,k) &= \lim_{\substack{q_i\to 0\\ q_j\to k}} \Delta\mathcal{F}_B^{(i,j)}(q_i,q_j)\\ &= -J_{ij} - \lim_{\substack{q_i\to 0\\ q_j\to k}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 0} - \overbrace{q_i}^{\to 0} - \overbrace{q_j}^{\to k}))\\ &= -J_{ij} + 2kJ_{ij}\\ &= J_{ij}\,(2k-1)\end{aligned}$$

and

$$\begin{aligned}\overline{\Delta\mathcal{F}_B^{(i,j)}}(k,0) &= -J_{ij} - \lim_{\substack{q_i\to k\\ q_j\to 0}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 0} - \overbrace{q_i}^{\to k} - \overbrace{q_j}^{\to 0}))\\ &= -J_{ij} + 2kJ_{ij}\\ &= J_{ij}\,(2k-1),\end{aligned}$$

with both expressions being monotonically increasing if k increases in $[0,1]$. On the other hand, it is monotonically decreasing over the two line segments that connect $(0,1)$ to $(1,1)$ and $(1,0)$ to $(1,1)$:

$$\begin{aligned}\overline{\Delta\mathcal{F}_B^{(i,j)}}(k,1) &= -J_{ij} - \lim_{\substack{q_i\to k\\ q_j\to 1}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 2k} - \overbrace{q_i}^{\to k} - \overbrace{q_j}^{\to 1}))\\ &= J_{ij} - 2kJ_{ij}\\ &= J_{ij}(1-2k)\end{aligned}$$

and

$$\begin{aligned}\overline{\Delta\mathcal{F}_B^{(i,j)}}(1,k) &= -J_{ij} - \lim_{\substack{q_i\to 1\\ q_j\to k}} 2J_{ij}\,(\overbrace{2\xi_{ij}}^{\to 2k} - \overbrace{q_i}^{\to 1} - \overbrace{q_j}^{\to k}))\\ &= J_{ij} - 2kJ_{ij}\\ &= J_{ij}(1-2k)\end{aligned}$$

with both expressions being monotonically decreasing if k increases in $[0,1]$. By this, we conclude that

$$\min_{(q_i,q_j)\in\overline{\mathbb{B}^{(i,j)}}} \overline{\Delta\mathcal{F}_B^{(i,j)}}(q_i,q_j) = -J_{ij}$$

and

$$\max_{(q_i,q_j)\in\overline{\mathbb{B}^{(i,j)}}} \overline{\Delta\mathcal{F}_B^{(i,j)}}(q_i,q_j) = J_{ij}.$$

According to the definition (15) of $\overline{\Delta\mathcal{F}_B^{(i,j)}}$ and in consequence of our previous observations, the statements (31), (32), and hence (30) follow immediately. □

PROOF OF COROLLARY 1:.

This is an immediate consequence of Theorem 1. □

PROOF OF THEOREM 2:.

We prove that the derivative of $||\Delta\mathcal{F}_B^{(i,j)}||_{L^2}^2$ with respect to J_{ij} is larger than 0 if $J_{ij} > 0$, and smaller than 0 if $J_{ij} < 0$. Recall that we assume $\mathcal{F}_B$ to be defined on the local polytope instead of the Bethe box (this implies also that $\Delta\mathcal{F}_B^{(i,j)}$ is defined on the sliced local polytope $\mathbb{L}^{(i,j)}$ 19). We compute

$$\frac{\partial}{\partial J_{ij}}||\Delta\mathcal{F}_B^{(i,j)}||_{L^2}^2 \tag{16}$$

$$= \frac{\partial}{\partial J_{ij}} \iiint_{\mathbb{L}^{(i,j)}} (\Delta\mathcal{F}_B^{(i,j)})^2 d\xi_{ij} dq_i dq_j \tag{17}$$

$$= \iiint_{\mathbb{L}^{(i,j)}} \frac{\partial}{\partial J_{ij}} (\Delta\mathcal{F}_B^{(i,j)})^2 d\xi_{ij} dq_i dq_j \tag{18}$$

$$= \iiint_{\mathbb{L}^{(i,j)}} 2 \cdot \Delta\mathcal{F}_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathcal{F}_B^{(i,j)} d\xi_{ij} dq_i dq_j \tag{19}$$

Now observe that $\frac{\partial}{\partial J_{ij}}\Delta\mathcal{F}_B^{(i,j)} = \frac{\partial}{\partial J_{ij}}(\Delta\mathcal{U}_B^{(i,j)} + I_B^{(i,j)})$ (this is just the definition of $\Delta\mathcal{F}_B^{(i,j)}$ (18)), and that the mutual information on the sliced local polytope is independent of J_{ij}. Consequently $\frac{\partial}{\partial J_{ij}} I_B^{(i,j)} = 0$. We continue with (19) and split the integral:

$$2 \iiint_{\mathbb{L}^{(i,j)}} (\Delta\mathcal{U}_B^{(i,j)} + I_B^{(i,j)}) \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathcal{U}_B^{(i,j)} d\xi_{ij} dq_i dq_j$$

$$= \underbrace{2 \iiint_{\mathbb{L}^{(i,j)}} \Delta\mathcal{U}_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathcal{U}_B^{(i,j)} d\xi_{ij} dq_i dq_j}_{(I.)} + \underbrace{2 \iiint_{\mathbb{L}^{(i,j)}} I_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathcal{U}_B^{(i,j)} d\xi_{ij} dq_i dq_j}_{(II.)}$$

The integral $(I.)$ can be simplified, as

$$\frac{\partial}{\partial J_{ij}} \Delta\mathcal{U}_B^{(i,j)} = -1 - 2\,(2\xi_{ij} - q_i - q_j) = \frac{1}{J_{ij}} \Delta\mathcal{U}_B^{(i,j)}$$

and therefore

$$(I.) = J_{ij} \iiint_{\mathbb{L}^{(i,j)}} (-1 - 2(2\xi_{ij} - q_i - q_j))^2 d\xi_{ij} dq_i dq_j,$$

which is larger than 0 for $J_{ij} > 0$ and smaller than 0 for $J_{ij} < 0$.

It remains to prove that the integral $(II.)$ is equal to zero. It is sufficient to consider the integral over the halfspace of the sliced local polytope where $q_i < q_j$ (due to symmetry, the integral over the other halfspace where $q_i > q_j$ is equal to the

first). We can split this halfspace in four 'sub-polytopes':

$$\begin{aligned}
\mathbb{L}_1^{(i,j)}: \ & \{(q_i,q_j,\xi_{ij}) \in \mathbb{L}^{(i,j)}: \\
& \quad 0 < q_j \leq \frac{1}{2}, 0 < q_i \leq q_j,\, 0 < \xi_{ij} < q_i\} \\
\mathbb{L}_2^{(i,j)}: \ & \{(q_i,q_j,\xi_{ij}) \in \mathbb{L}^{(i,j)}: \\
& \quad \frac{1}{2} \leq q_j < 1, 0 < q_i \leq 1-q_j,\, 0 < \xi_{ij} < q_i\} \\
\mathbb{L}_3^{(i,j)}: \ & \{(q_i,q_j,\xi_{ij}) \in \mathbb{L}^{(i,j)}: \\
& \quad \frac{1}{2} \leq q_j < 1, 1-q_j \leq q_i \leq \frac{1}{2}, \\
& \quad q_i + q_j - 1 < \xi_{ij} < q_i\} \\
\mathbb{L}_4^{(i,j)}: \ & \{(q_i,q_j,\xi_{ij}) \in \mathbb{L}^{(i,j)}: \\
& \quad \frac{1}{2} \leq q_j < 1, \frac{1}{2} \leq q_i \leq q_j, \\
& \quad q_i + q_j - 1 < \xi_{ij} < q_i\}
\end{aligned} \tag{20}$$

We show that, for each point $(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)})$ in $\mathbb{L}_1^{(i,j)}$ we can find a point $(q_i^{(2)}, q_j^{(2)}, \xi_{ij}^{(2)})$ in $\mathbb{L}_2^{(i,j)}$ such that the value of the integrand $I_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathfrak{U}_B^{(i,j)}$ of $(II.)$ in $(q_i^{(2)}, q_j^{(2)}, \xi_{ij}^{(2)})$ is precisely the negative of the value that $I_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathfrak{U}_B^{(i,j)}$ takes in $(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)})$. To this end, we map $(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)})$ from $\mathbb{L}_1^{(i,j)}$ to $(q_i^{(1)}, 1-q_j^{(1)}, q_i^{(1)} - \xi_{ij}^{(1)})$ in $\mathbb{L}_2^{(i,j)}$.

Then one can check that

$$\begin{aligned}
& \Delta\mathfrak{U}_B^{(i,j)}(q_i^{(1)}, 1-q_j^{(1)}, q_i^{(1)} - \xi_{ij}^{(1)}) \\
= & -\Delta\mathfrak{U}_B^{(i,j)}(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)})
\end{aligned}$$

and

$$\begin{aligned}
& I_B^{(i,j)}(q_i^{(1)}, 1-q_j^{(1)}, q_i^{(1)} - \xi_{ij}^{(1)}) \\
= & I_B^{(i,j)}(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)}),
\end{aligned}$$

and consequently

$$\begin{aligned}
& I_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathfrak{U}_B^{(i,j)}(q_i^{(1)}, 1-q_j^{(1)}, q_i^{(1)} - \xi_{ij}^{(1)}) \\
= & -I_B^{(i,j)} \cdot \frac{\partial}{\partial J_{ij}} \Delta\mathfrak{U}_B^{(i,j)}(q_i^{(1)}, q_j^{(1)}, \xi_{ij}^{(1)}).
\end{aligned}$$

As our so defined mapping is bijective, we conclude that the integral $(II.)$ over $\mathbb{L}_2^{(i,j)}$ is exactly the negative of the integral $(I.)$ over $\mathbb{L}_1^{(i,j)}$. Similarly, we can define a bijective mapping between $\mathbb{L}_3^{(i,j)}$ and $\mathbb{L}_4^{(i,j)}$ with the same result. We summarize, that the integral $(II.)$ over the entire sliced local polytope equals to zero, which completes the proof. □

PROOF OF THEOREM 3:.

If $\mathcal{Z}_\mathcal{B} = \mathcal{Z}_\mathcal{B}{}^{\setminus(i,j)}$, the statement is trivial and we will therefore exclude this scenario. According to Theorem 1, each edge contributes less than $\pm J_{ij}$ to the 'overall' Bethe function and removing an edge can only transpose the individual function values by less than $|J_{ij}|$. This means that for all $\boldsymbol{q} \in \mathbb{B}$,

$$|\mathcal{F}_B(\boldsymbol{q}) - \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q})| = |\Delta\mathcal{F}_B^{(i,j)}(q_i, q_j)| < |J_{ij}|. \tag{21}$$

In particular, this is also valid for the global minimizer $\boldsymbol{q_1^*}$ of $\mathcal{F}_B$ as well as for the global minimizer $\boldsymbol{q_2^*}$ of $\mathcal{F}_B^{\setminus(i,j)}$. We now distinguish between two cases:

Case 1: $\mathcal{Z}_\mathcal{B} > \mathcal{Z}_\mathcal{B}{}^{\setminus(i,j)}$, or equivalently, $\mathcal{F}_B(\boldsymbol{q_1^*}) < \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*})$. Then

$$0 < \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*}) - \mathcal{F}_B(\boldsymbol{q_1^*}) < \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_1^*}) - \mathcal{F}_B(\boldsymbol{q_1^*}) \overset{(21)}{<} |J_{ij}|.$$

Case 2: $\mathcal{Z}_{\mathcal{B}} < {\mathcal{Z}_{\mathcal{B}}}^{\setminus(i,j)}$, or equivalently, $\mathcal{F}_B(\boldsymbol{q_1^*}) > \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*})$. Then

$$0 < \mathcal{F}_B(\boldsymbol{q_1^*}) - \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*}) < \mathcal{F}_B(\boldsymbol{q_2^*}) - \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*}) \overset{(21)}{<} |J_{ij}|.$$

Together, we finally obtain

$$\begin{aligned} |J_{ij}| &> |\mathcal{F}_B(\boldsymbol{q_1^*}) - \mathcal{F}_B^{\setminus(i,j)}(\boldsymbol{q_2^*})| \\ &= |\min_{\boldsymbol{q}\in\mathbb{B}} \mathcal{F}_B - \min_{\boldsymbol{q}\in\mathbb{B}} \mathcal{F}_B^{\setminus(i,j)}| = \\ &= |\log(\mathcal{Z}_{\mathcal{B}}) - \log({\mathcal{Z}_{\mathcal{B}}}^{\setminus(i,j)})| \\ &= \left| \log\Big(\frac{\mathcal{Z}_{\mathcal{B}}}{{\mathcal{Z}_{\mathcal{B}}}^{\setminus(i,j)}}\Big)\right|. \end{aligned}$$

□

PROOF OF THEOREM 4:.

Without loss of generality, we assume that all θ_i are larger than 0.5. From Lemma 4 [Knoll et al., 2021] we know, that the global minimum $-\log(\mathcal{Z}_{\mathcal{B}})$ of $\mathcal{F}_B$ must be located in a point of $\mathbb{B}$ where all components q_i are larger than 0.5. If we remove an edge, we obtain again a unidirectional model that satisfies the same property. By Lemma 6, we then know that the global minimum of $\mathcal{F}_B$ in the new model must be larger than the global minimum in the old model, i.e.,

$$-\log(\mathcal{Z}_{\mathcal{B}}) < -\log({\mathcal{Z}_{\mathcal{B}}}^{\setminus(i,j)}) \tag{22}$$

(as the removed energy difference function $\Delta\mathcal{F}_B^{(i,j)}$ makes in the original model a negative contribution to the energy of all points $\boldsymbol{q} \in \mathbb{B}$ with $q_i, q_j > 0.5$ – in particular, to the global minimum of the new model). Finally, we apply Theorem 2 from Ruozzi [2012] which says that in attractive models, the Bethe partition function is always a lower bound to the true partition function, i.e., $-\log(\mathcal{Z}) < -\log(\mathcal{Z}_{\mathcal{B}})$. Together with (22), we conclude

$$|-\log(\mathcal{Z}) + \log(\mathcal{Z}_{\mathcal{B}})| < |-\log(\mathcal{Z}) + \log({\mathcal{Z}_{\mathcal{B}}}^{\setminus(i,j)})|,$$

which is equivalent to the statement. □

APPENDIX B – RESULTS FROM RELATED WORK

Lemma 1. *Let X, Y discrete random variables with alphabets $\mathcal{X}, \mathcal{Y}$. Their mutual information $I(X;Y)$ is bounded by*

$$\frac{1}{2}\Big(\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} |p_{XY}(x,y) - p_X(x)p_Y(y)|\Big)^2 \;\leq\; I(X;Y) \;\leq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{p_{XY}^2(x,y)}{p_X(x)p_Y(y)} \;-\; 1. \tag{23}$$

Proof. Corollary 1 and 2 in Dragomir et al. [2000]. □

Lemma 2.

(a) If (i,j) is an attractive edge (i.e., $J_{ij} > 0$), then $\xi_{ij}^* > q_i q_j$.

(b) If (i,j) is an repulsive edge (i.e., $J_{ij} < 0$), then $\xi_{ij}^* < q_i q_j$.

Proof. Lemma 2 in Weller and Jebara [2013]. □

Lemma 3. *The first-order derivatives of the Bethe free energy $\mathcal{F}_B$ on the Bethe box $\mathbb{B}$ are*

$$\frac{\partial}{\partial q_i}\mathcal{F}_B = -2\,\theta_i + 2\sum_{j\in\mathcal{N}(i)} J_{ij} + \log\Big(\frac{(1-q_i)^{d_i-1}}{q_i^{d_i-1}} \prod_{j\in\mathcal{N}(i)} \frac{q_i - \xi_{ij}^*}{1 + \xi_{ij}^* - q_i - q_j}\Big). \tag{24}$$

Proof. This is an intermediate result in Welling and Teh [2001]. □

Theorem 1. *The second partial derivatives of edge specific Bethe terms of the form*

$$f_{ij}(q_i, q_j) = -\left(1 + 2\left(2\,\xi_{ij} - q_i - q_j\right)\right) J_{ij} - \mathcal{S}_{ij} \tag{25}$$

are calculated as

$$\frac{\partial^2}{\partial q_i^2} f_{ij} = \frac{q_j(1-q_j)}{T_{ij}}, \tag{26}$$

$$\frac{\partial^2}{\partial q_i q_j} f_{ij} = \frac{\partial^2}{\partial q_j q_i} f_{ij} = \frac{q_i q_j - \xi_{ij}^*}{T_{ij}}, \quad \textit{and} \tag{27}$$

$$\frac{\partial^2}{\partial q_j^2} f_{ij} = \frac{q_i(1-q_i)}{T_{ij}}, \tag{28}$$

where $T_{ij} \coloneqq q_i q_j (1-q_i)(1-q_j) - (\xi_{ij}^* - q_i q_j)^2$.

Proof. Theorem 10 in Weller and Jebara [2013]. □

Lemma 4. *Consider a unidirectional model and assume without loss of generality that* $\theta_i > 0.5$*. Then, in the global minimum* $\boldsymbol{q}^*$ *of* $\mathcal{F}_B$ *all components* q_i *are* > 0.5*.*

Proof. Lemma 2 in Knoll et al. [2021]. □

Theorem 2. *For an attractive model with binary variables, the Bethe partition function is always a lower bound on the true partition function. That is,*

$$\mathcal{Z}_{\mathcal{B}} \leq \mathcal{Z},$$

or equivalently,

$$\min_{\mathbb{M}} \mathcal{F} = -\log(\mathcal{Z}) \leq -\log(\mathcal{Z}_{\mathcal{B}}) = \min_{\mathbb{L}} \mathcal{F}_B.$$

Proof. Theorem 4.1 in Ruozzi [2012]. □

APPENDIX C – FURTHER EXPERIMENTS

Here we present supplementary experiments that we have performed on a (non-toroidal) 5×5-grid graph. The experimental setting is the same as in Sec. 4. Again, we observe the beneficial effects of edge removal on the approximated marginals. In this sparser model, the treshold beyond which the Bethe free energy becomes non-convex and the accuracy of BP degrades is larger than for the fully connected graph considered in Sec. 4. Beyond that treshold, edge removal is particularly effective if the local potentials are weak; but also for models with strong local potentials the results are mostly superior to the marginal accuracy of BP in the original model. For attractive models with strong local potentials, we observe again the presence of a 'channel' that specifies an optimal intermediate model state.

References

Sever S. Dragomir, Marcel L. Scholz, and Jadranka Sunde. Some upper bounds for the relative entropy and applications. *Computers & Mathematics with Applications*, 39(9–10):91–100, 2000.

Christian Knoll, Adrian Weller, and Franz Pernkopf. Self-guided belief propagation – a homotopy continuation method. In *arXiv:1812.01339*, 2021.

Nicholas Ruozzi. The Bethe partition function of log-supermodular graphical models. In *Proceedings of NIPS*, 2012.

Adrian Weller and Tony Jebara. Bethe bounds and approximating the global optimum. In *Proceedings of AISTATS*, 2013.

Max Welling and Yee W. Teh. Belief optimization for binary networks: A stable alternative to loopy belief propagation. In *Proceedings of UAI*, 2001.

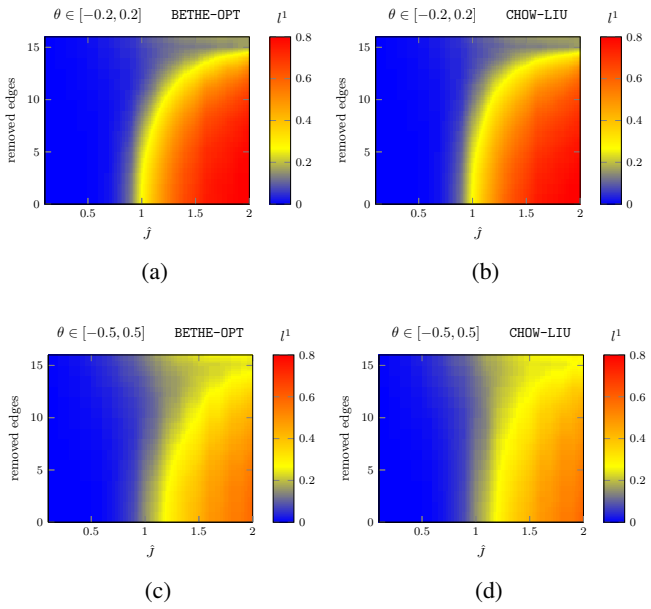

Figure 1: Attractive models (5x5 - grid graph). First row: $\theta_i \in [-0.2, 0.2]$; second row: $\theta_i \in [-0.5, 0.5]$. (a) + (c): `BETHE-OPT` criterion; (b) + (d): `CHOW LIU` criterion.

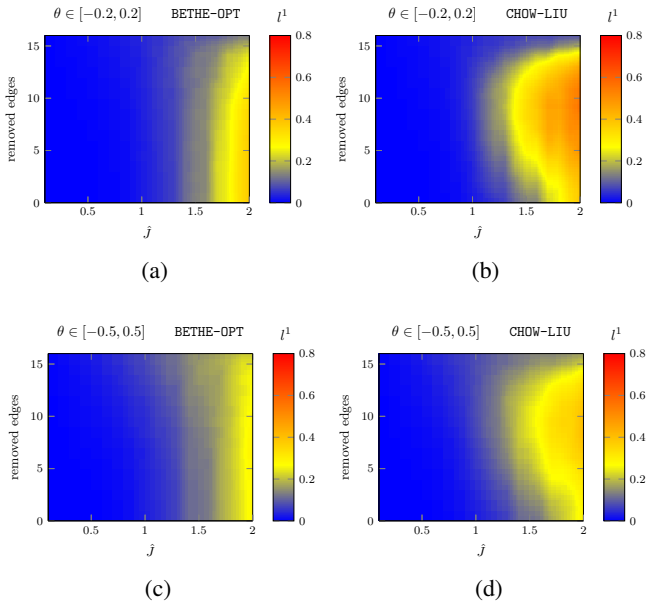

Figure 2: General models (5x5 - grid graph). First row: $\theta_i \in [-0.2, 0.2]$; second row: $\theta_i \in [-0.5, 0.5]$. (a) + (c): `BETHE-OPT` criterion; (b) + (d): `CHOW LIU` criterion.