APPENDIX: SUPPLEMENTARY MATERIAL
Uprooting and Rerooting Graphical Models

Here we provide:

- Proof of Lemma 2.
- In §8, details of the polytopes and proofs from §5.2, including proofs of Theorems 3 and 4.
- In §9, details of experimental methods, and additional results.

Lemma 2. Given distribution p_i for any $i \in \{1, \ldots, n\}$, the marginals $p_0(X_j = 1)$ for the original model $M = M_0$ may be recovered as follows:

$$p_0(X_j = 1) = \begin{cases} p_i(X_0 = 1) & j = i \\ p_i(X_j \neq X_0) & j \neq i. \end{cases}$$

Proof. First, for $j = i$ we have

$$\begin{aligned} p_0(X_i = 1) &= p_+(X_i = 1 | X_0 = 0) \\ &= \frac{p_+(X_i = 1, X_0 = 0)}{p_+(X_0 = 0)} \\ &= \frac{p_+(X_i = 0, X_0 = 1)}{p_+(X_i = 0)} \quad \text{(symmetry of } M^+, \text{ note that } p_+(X_r = 0) = \frac{1}{2} \text{ for any } r \in \{0, \ldots, n\}) \\ &= p_i(X_0 = 1). \end{aligned}$$

Next, for $j \neq i$, again using symmetry of M^+,

$$\begin{aligned} p_0(X_j = 1) &= p_+(X_j = 1 | X_0 = 0) \\ &= \frac{p_+(X_j = 1, X_0 = 0)}{p_+(X_0 = 0)} \\ &= \frac{p_+(X_j = 1, X_0 = 0, X_i = 0) + p_+(X_j = 1, X_0 = 0, X_i = 1)}{p_+(X_0 = 0)} \\ &= \frac{p_+(X_j = 1, X_0 = 0, X_i = 0) + p_+(X_j = 0, X_0 = 1, X_i = 0)}{p_+(X_i = 0)} \\ &= \frac{p_+(X_j \neq X_0, X_i = 0)}{p_+(X_i = 0)} \\ &= p_i(X_j \neq X_0). \end{aligned}$$

□

8. Details of the polytopes and proofs from section 5.2

Weller et al. (2016) showed that LP+TRI is tight (that is, the LP relaxation on the triplet-consistent polytope is guaranteed to yield an optimum at an integral vertex) for any model which is almost balanced (that is, any model which contains a variable s.t. if it is removed then the remaining model is balanced; any singleton potentials are allowed). We first provide background and preliminary results in §8.1-8.2. For more extensive background, see (Wainwright and Jordan, 2008, Chapter 8), (Sontag, 2007) or (Deza and Laurent, 1997).

In §8.3, we prove Theorem 3, a general result which shows that TRI is 'universally rooted'. Many optimization results that apply for TRI for *some* rerooting of a model will automatically apply *for all* rerootings.

We shall apply Theorem 3 to show how the result of Weller et al. (2016) may be significantly strengthened in Theorem 4 to demonstrate tightness of LP+TRI for any model M whose uprooted model M^+ is 2-almost balanced (that is, the uprooted model contains 2 variables s.t. if they are both removed then what remains in the uprooted model is balanced).

Notation. As in 4.1.2, in order to differentiate between probabilities obtained for an initial model $M = M_0$, its uprooted model M^+, and various rerooted models M_i, we use the following notation: let p_i be the probability distribution in model M_i, in particular p_0 is the distribution for model M_0 which is the original model M; let p_+ be the distribution in the uprooted model M^+.

Using similar reasoning to that used above in the proof of Lemma 2, we use the symmetry of M^+ to show the following results which will be useful in §8.3 for mapping rooted probabilities p_i to 'universal' uprooted probabilities p_+.

Lemma 5. *(i) For any distinct $i, j \in \{0, \dots, n\}$, $p_i(X_j = 1) = p_+(X_i \neq X_j)$;*
(ii) for any distinct $i, j, k \in \{0, \dots, n\}$, $p_i(X_j \neq X_k) = p_+(X_j \neq X_k)$.

Proof. (i) For distinct $i, j \in \{0, \dots, n\}$,

$$\begin{aligned} p_i(X_j = 1) &= p_+(X_j = 1 | X_i = 0) \\ &= \frac{p_+(X_j = 1, X_i = 0)}{p_+(X_i = 0)} \\ &= 2p_+(X_j = 1, X_i = 0) \\ &= p_+(X_j = 1, X_i = 0) + p_+(X_j = 0, X_i = 1) \quad \text{(symmetry of } M^+\text{)} \\ &= p_+(X_i \neq X_j). \end{aligned}$$

(ii) For distinct $i, j, k \in \{0, \dots, n\}$,

$$\begin{aligned} p_i(X_j \neq X_k) &= p_+(X_j \neq X_k | X_i = 0) \\ &= \frac{p_+(X_j = 1, X_k = 0, X_i = 0) + p_+(X_j = 0, X_k = 1, X_i = 0)}{p_+(X_i = 0)} \\ &= 2\left[p_+(X_j = 1, X_k = 0, X_i = 0) + p_+(X_j = 0, X_k = 1, X_i = 0)\right] \\ &= p_+(X_j = 1, X_k = 0, X_i = 0) + p_+(X_j = 0, X_k = 1, X_i = 0) \\ &\quad + p_+(X_j = 0, X_k = 1, X_i = 1) + p_+(X_j = 1, X_k = 0, X_i = 1) \quad \text{(symmetry of } M^+\text{)} \\ &= p_+(X_j = 1, X_k = 0) + p_+(X_j = 0, X_k = 1) \\ &= p_+(X_j \neq X_k). \end{aligned}$$

□

8.1. The marginal polytope and its relaxations LOC and TRI

Given a model M with n variables $\mathcal{V}$ and m edges $\mathcal{E}$, we may consider a vector containing marginal probabilities for all n single variables and all m pairs of variables that are directly related.

Specifically, regarding the score (1), for any configuration $x = (x_1, \dots, x_n)$, let $y_{ij} = \mathbb{1}[x_i \neq x_j]$ then collect the x and y terms together into a vector $z = (x_1, \dots, x_n, \dots, y_{ij}, \dots) \in \{0, 1\}^{n+m}$. Similarly collect together the potential parameters into a vector $w = (\theta_1, \dots, \theta_n, \dots, -\frac{1}{2}W_{ij}, \dots) \in \mathbb{R}^{n+m}$. Now the score of a configuration x may be written as $w \cdot z(x)$, and MAP inference may be framed as an integer linear program to find $z^* \in \arg\max_{z : x \in \{0,1\}^n} w \cdot z$.

The convex hull of the 2^n possible integer solutions in $[0, 1]^{n+m}$ is the *marginal polytope* $\mathbb{M}$ for our choice of singleton and edge terms in (1). Regarding the convex coefficients as a probability distribution p_0 over all possible states, the marginal polytope may be considered the space of all singleton and pairwise mean marginals that are consistent with some global distribution p_0 over the 2^n states, that is

$$\mathbb{M} = \{\mu = (\mu_1, \dots, \mu_n, \dots, \mu_{ij}, \dots) \in [0, 1]^d \text{ s.t. } \exists p_0 : \mu_i = \mathbb{E}_{p_0}(X_i)\ \forall i, \mu_{ij} = \mathbb{E}_{p_0}(\mathbb{1}[X_i \neq X_j])\ \forall (i, j) \in E\}. \quad (4)$$

Note that $\mu_i = p_0(X_i = 1)$ and $\mu_{ij} = p_0(X_i \neq X_j)$.

Since an LP attains an optimum at a vertex of the feasible region, if $w \cdot \mu$ is maximized over $\mathbb{M}$ then an exact integer solution is always optimum. However, $\mathbb{M}$ has exponentially many facets (Deza and Laurent, 1997), hence a simpler, relaxed constraint set is typically employed, yielding an upper bound on the original optimum. This set is often chosen as the *local polytope* LOC, which enforces only pairwise consistency (Wainwright and Jordan, 2008). If an optimum vertex is achieved at an integer solution, then this must be an optimum of the original discrete problem, in which case we say

that the relaxation LP+LOC is *tight*. Sherali and Adams (1990) proposed a series of successively tighter relaxations by enforcing consistency over progressively larger clusters of variables. At order r, the $\mathcal{L}_r$ polytope enforces consistency over all clusters of variables of size $\leq r$. $\mathcal{L}_2$ is the local polytope LOC. Next, $\mathcal{L}_3$ is the triplet-consistent polytope TRI, and so on, with $\mathcal{L}_n = \mathbb{M} \subseteq \mathcal{L}_{n-1} \subseteq \cdots \subseteq \mathcal{L}_3 = \text{TRI} \subseteq \mathcal{L}_2 = \text{LOC}$.

In order to obtain the explicit constraints for these polytopes, earlier work (Wainwright and Jordan, 2008; Weller et al., 2016) uses a different (but equivalent) minimal reparameterization leading to a different (but equivalent) set of marginals. To link to their notation, let $\alpha_i = p_0(X_i = 1), \alpha_{ij} = p_0(X_i = 1, X_j = 1), \alpha_{ijk} = p_0(X_i = 1, X_j = 1, X_k = 1)$. We next present a derivation of the constraints for LOC and TRI following (Weller et al., 2016), see also (Wainwright and Jordan, 2008, Example 8.7).

Examining just one variable, we have $\alpha_i = \mu_i \in [0, 1]\ \forall i$. In order to be consistent with these single variable marginals, the matrix of pairwise marginals for edge (i, j) takes the form

$$\begin{pmatrix} p_0(X_i = 0, X_j = 0) & p_0(X_i = 0, X_j = 1) \\ p_0(X_i = 1, X_j = 0) & p_0(X_i = 1, X_j = 1) \end{pmatrix} = \begin{pmatrix} 1 + \alpha_{ij} - \alpha_i - \alpha_j & \alpha_j - \alpha_{ij} \\ \alpha_i - \alpha_{ij} & \alpha_{ij} \end{pmatrix}. \tag{5}$$

The LOC constraints are exactly those that ensure that all 4 terms are ≥ 0, which leads to

$$\text{LOC constraints for edge } (i, j): \qquad \max(0, \alpha_i + \alpha_j - 1) \leq \alpha_{ij} \leq \min(\alpha_i, \alpha_j). \tag{6}$$

These constraints may be reformulated in terms of our μ_{ij} marginals by using $\mu_i = \alpha_i$, and observing from (5) that

$$\mu_{ij} = \alpha_i + \alpha_j - 2\alpha_{ij} \quad \Leftrightarrow \quad \alpha_{ij} = \frac{1}{2}\left(\mu_i + \mu_j - \mu_{ij}\right).^{5} \tag{7}$$

To obtain the constraints for TRI, we use a 'lift-and-project' approach by 'lifting' to distributions over three variables, deriving conditions, then projecting these back down to the one and two variable marginals that we are using. We must ensure that the distribution over every triplet of variables X_i, X_j, X_k is valid and consistent with all edge and singleton marginals. Given $\alpha_i, \alpha_j, \alpha_k, \alpha_{ij}, \alpha_{ik}, \alpha_{jk}$ and using $\alpha_{ijk} = p_0(X_i = 1, X_j = 1, X_k = 1)$ as defined above, we have:

With $k = 0$,

$$\begin{pmatrix} p_0(X_i = 0, X_j = 0) & p_0(X_i = 0, X_j = 1) \\ p_0(X_i = 1, X_j = 0) & p_0(X_i = 1, X_j = 1) \end{pmatrix} = \begin{pmatrix} 1 - \alpha_i - \alpha_j - \alpha_k + \alpha_{ij} + \alpha_{ik} + \alpha_{jk} - \alpha_{ijk} & \alpha_j + \alpha_{ijk} - \alpha_{ij} - \alpha_{jk} \\ \alpha_i + \alpha_{ijk} - \alpha_{ij} - \alpha_{ik} & \alpha_{ij} - \alpha_{ijk} \end{pmatrix}$$

With $k = 1$,

$$\begin{pmatrix} p_0(X_i = 0, X_j = 0) & p_0(X_i = 0, X_j = 1) \\ p_0(X_i = 1, X_j = 0) & p_0(X_i = 1, X_j = 1) \end{pmatrix} = \begin{pmatrix} \alpha_k + \alpha_{ijk} - \alpha_{ik} - \alpha_{jk} & \alpha_{jk} - \alpha_{ijk} \\ \alpha_{ik} - \alpha_{ijk} & \alpha_{ijk} \end{pmatrix}.$$

As previously for LOC, we have the constraints that all terms are ≥ 0. By combining inequalities, we may project back down by eliminating α_{ijk}. For example, if we combine the condition that the top right element of the matrix for $k = 0$ is ≥ 0 with the similar condition for the bottom right element of the same matrix, we obtain $\alpha_j - \alpha_{jk} \geq 0$ which is one of the LOC constraints for edge (j, k), see (6). Working through the various combinations yields all the previous LOC constraints for the edges $(i, j), (i, k)$ and (j, k), and in addition we obtain the following four new *triplet constraints*, which are called *cycle inequalities* in (Wainwright and Jordan, 2008, Example 8.7).

$$\begin{aligned} \text{TRI constraints in terms of } \alpha \text{ marginals for triplet of distinct } i, j, k \in \{1, \ldots, n\}: \qquad \alpha_i + \alpha_{jk} &\geq \alpha_{ij} + \alpha_{ik} \\ \alpha_j + \alpha_{ik} &\geq \alpha_{ij} + \alpha_{jk} \\ \alpha_k + \alpha_{ij} &\geq \alpha_{ik} + \alpha_{jk} \\ \alpha_{ij} + \alpha_{ik} + \alpha_{jk} &\geq \alpha_i + \alpha_j + \alpha_k - 1. \end{aligned} \tag{8}$$

[5] This equivalence is essentially the *covariance mapping* described in (Deza and Laurent, 1997, §5.2).

If we use (7) to rewrite these TRI constraints (8) in terms of μ marginals, then they take the following appealing form.

$$\text{TRI constraints in terms of } \mu \text{ marginals for triplet of distinct } i,j,k \in \{1,\dots,n\}: \quad \begin{aligned} \mu_{jk} &\le \mu_{ij} + \mu_{ik} \\ \mu_{ik} &\le \mu_{ij} + \mu_{jk} \\ \mu_{ij} &\le \mu_{ik} + \mu_{jk} \\ \mu_{ij} + \mu_{ik} + \mu_{jk} &\le 2. \end{aligned} \tag{9}$$

Notice that (9) considers only terms of the form μ_{ij}. Since μ_{ij} is the probability that X_i and X_j take different values, a simple way to see that the inequalities of (11) are valid is to observe that they clearly hold for any integer settings of $X_i, X_j, X_k \in \{0,1\}^3$, and hence they must hold for any valid probability distribution over the 8 possible settings of these three variables (since this yields a convex combination).

8.2. The cut polytope and its relaxations RMET and MET

As in §3: given a model M with variables $\{X_1,\dots,X_n\}$ on graph $G(\mathcal{V},\mathcal{E})$ with vertices $\mathcal{V} = \{1,\dots,n\}$ and edges $\mathcal{E}$, its uprooted model M^+ has variables $\{X_0,\dots,X_n\}$ on graph $G'(\mathcal{V}',\mathcal{E}')$ with vertices $\mathcal{V}' = \{0,1,\dots,n\}$ and edges $\mathcal{E}' = \mathcal{E} \cup \mathcal{F}$, where $\mathcal{F} = \{(0,i) : \theta_i \neq 0\}$. An uprooted model M^+ is completely symmetric. The score (2) considers only edges and examines only whether the end variables of each edge take the same value.

Given a subset $S \subseteq \mathcal{V}' = \{0,1,\dots,n\}$, let $\delta(S) \in \{0,1\}^{|\mathcal{E}'|}$ be the *cut vector* of edges of $\mathcal{E}'$ which run between the vertex partitions S and $\mathcal{V}' \setminus S$, defined by $\delta(S)_{ij} = 1$ iff i and j are in different partitions.

The *cut polytope* (Barahona and Mahjoub, 1986) of G' is the convex hull of all such cut vectors, that is CUT = conv $\{\delta(S) : S \subseteq \mathcal{V}'\}$. Although there are 2^{n+1} choices of S, CUT has 2^n vertices since by definition $\delta(S) = \delta(\mathcal{V}' \setminus S)$. In fact, there is a simple linear bijection between CUT and the marginal polytope $\mathbb{M}$ of M.

Given $d \in$ CUT with entries $d(i,j)$ for each edge $(i,j) \in \mathcal{E}'$, d maps to $\mu \in \mathbb{M}$ where $\mu_j = d(0,j)$ for $j \in \mathcal{V}$, and $\mu_{ij} = d(i,j)$ for $(i,j) \in \mathcal{E}$. To see this, $d(i,j)$ may be interpreted as the marginal probability that $i,j \in \mathcal{V}'$ lie in different partitions.

As an aside, note that the marginal polytope of M^+, which we call $\mathbb{M}^+$, is closely related, but different, to CUT. $\mathbb{M}^+$ has $n+1$ additional dimensions for the singleton marginal dimensions of its $n+1$ variables, though given the symmetry of M^+, these are all 1/2.

MAP inference for the model M on G is equivalent to the weighted max cut problem for G':

$$\max_{\mu \in \mathbb{M}} w \cdot \mu = \max_{e \in \text{CUT}} w' \cdot d, \quad w'_{ij} = \begin{cases} \theta_j & i = 0 \\ -\frac{1}{2} W_{ij} & (i,j) \in E. \end{cases} \tag{10}$$

The bijection between $\mathbb{M}$ and CUT may also be used to map the LOC and TRI relaxations of $\mathbb{M}$ to corresponding relaxations of CUT in $[0,1]^{|E'|}$, called the *rooted semimetric polytope* RMET and the *semimetric polytope* MET, respectively. The constraints for the MET polytope (which corresponds to TRI) take the following form, sometimes described as unrooted triangle inequalities (Deza and Laurent, 1997, §27.1):

$$\text{MET constraints } \forall \text{ distinct } i,j,k \in V' = \{0,1,\dots,n\}: \quad \begin{aligned} d(i,j) - d(i,k) - d(j,k) &\le 0 \\ d(i,j) + d(i,k) + d(j,k) &\le 2. \end{aligned} \tag{11}$$

Note that the MET constraints (11) restricted to triplets $i,j,k \in \mathcal{V} = \{1,\dots,n\}$ are identical to the TRI constraints for μ marginals in (9). Both enforce triplet consistency on the marginal probabilities of edges having end vertices which are different.

Remarkably, the constraints on d for RMET, the *rooted* triangle inequalities, which are equivalent to the LOC constraints on μ for LOC (6), are exactly just those of (11) for which one of i,j,k is 0, the vertex that was added to G to yield G'. Hence, RMET may be regarded as MET *rooted* at 0. Correspondingly, we may consider TRI to be a version of LOC that is *universally rooted*.

To see this, we shall consider the LOC constraints for edge $(i,j) \in \mathcal{E}$ (6), and show that they are exactly the MET constraints (11) applied to triplet $(0,i,j)$ in $\mathcal{V}'$. Consider the triangle $0ij$ of G' shown in Figure 4.

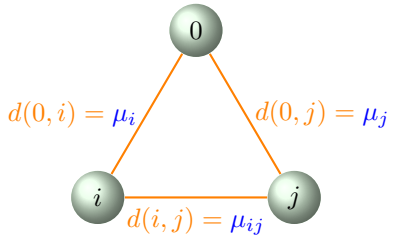

Figure 4. Illustration of edge marginals for the MET polytope, shown in orange, and their values in terms of μ marginals for the model M, shown in blue.

Recall that $\mu_i = \alpha_i$, $\mu_j = \alpha_j$, and from (7) that $\mu_{ij} = \alpha_i + \alpha_j - 2\alpha_{ij}$. Hence, the MET constraints (11) with $k = 0$ become:

$$\begin{aligned}
d(i,j) - d(i,k) - d(j,k) \le 0 \quad &\Leftrightarrow \quad \alpha_i + \alpha_j - 2\alpha_{ij} - \alpha_i - \alpha_j \le 0 \quad \Leftrightarrow \quad \alpha_{ij} \ge 0 \\
d(i,k) - d(i,j) - d(j,k) \le 0 \quad &\Leftrightarrow \quad \alpha_i - \alpha_i - \alpha_j + 2\alpha_{ij} - \alpha_j \le 0 \quad \Leftrightarrow \quad \alpha_{ij} \le \alpha_j \\
d(j,k) - d(i,j) - d(i,k) \le 0 \quad &\Leftrightarrow \quad \alpha_j - \alpha_i - \alpha_j + 2\alpha_{ij} - \alpha_i \le 0 \quad \Leftrightarrow \quad \alpha_{ij} \le \alpha_i \\
d(i,j) + d(i,k) + d(j,k) \le 2 \quad &\Leftrightarrow \quad \alpha_i + \alpha_j - 2\alpha_{ij} + \alpha_i + \alpha_j \le 2 \quad \Leftrightarrow \quad \alpha_{ij} \ge \alpha_i + \alpha_j - 1,
\end{aligned}$$

which exactly match the LOC constraints (6), as required.

8.3. New results

With the background in §8.1-8.2, we are ready to prove our new results.

Notation. Let μ^i, w^i be the μ, w vectors corresponding to rerootings at X_i. In particular, μ^0, w^0 are the μ, w vectors for the original model $M = M_0$.

Theorem 3. (TRI is 'universally rooted') LP+TRI yields the same optimum score for M as for any rerooting M_i; hence LP+TRI is either tight for all rerootings or for none.

Proof. First, note that the MAP score for M is the same as that for any rerooting M_i. One way to see this follows the observations in §3-4: each configuration x of M maps to exactly 2 configurations of M^+: $y_0 = (0, x)$ and $y_1 = \bar{y}_0 = (1, \bar{x})$, with the potentials of M^+ set so that score(x) = score(y_0) = score(y_1). Hence, in particular, a MAP configuration for M maps to two MAP configurations for M^+ with the same score, and exactly one of these will be in any rerooting M_i as a MAP configuration for that rerooted model with the same score.

It remains to show that $\max_{\text{TRI}(M_i)} w^i \cdot \mu^i$ is the same for any rerooting of a model M. We shall use a similar idea, converting the problem for M into a problem over the graph G' of the uprooted model, in such a way that this problem over G' is the same for all rerootings M_i. In fact, we shall show a score-preserving linear bijection between TRI(M_i) and MET, where we must still show that this is the same no matter which rerooting M_i is used.

In §8.2, we gave a simple linear bijection between $\mathbb{M}$ and CUT, which naturally extends to a linear bijection between TRI(M) and MET(M). Further it is clear that this is score preserving if we use w' from (10). That is, we have for any $\mu \in$ TRI(M), a linear bijection between μ and $d \in$ MET(M) s.t. $w \cdot \mu = w' \cdot d$. If these are maximized over their respective (equivalent) polytopes, then we obtain the same maximum.

It remains to show that for all rerootings, MET(M) = MET(M_i) and that w^i maps to the same vector w' for each MET. MET(M) = MET(M_i) follows directly from Lemma 5. Each w^i maps to the same vector w' by construction, see (2). □

The next result follows as a simple application of Theorem 3 to the earlier result of Weller et al. (2016).

Theorem 4. LP+TRI is tight for (any rerooting of) a model M whose uprooted model M^+ is 2-almost balanced.

Proof. First, if M^+ is 2-almost balanced with special variables X_i and X_j, then if we root at either X_i or X_j, we obtain an almost balanced model (that is, M_i or M_j) on which LP+TRI is tight by the result of Weller et al. (2016). Now if LP+TRI is tight for M_i, then by Theorem 3, LP+TRI is tight for any rerooting of M_i, including M. □

Following our result, Weller (2016) demonstrated a still stronger result: LP+TRI is tight for any model M whose uprooted model M^+ does not contain an *odd-*K_5 as a *signed minor*. An odd-K_5 is the complete graph on 5 variables where all edges are repulsive. Since an odd-K_5 is clearly not 2-almost balanced (if any 2 variables are removed, the remaining model is a frustrated triangle), all 2-almost balanced models are a subset of those that do not contain an odd-K_5 as a signed minor. Further, the condition of Weller (2016) was shown to be both sufficient and necessary for tightness for models with all potentials that respect the edge signs of the uprooted model. For details, see (Weller, 2016).

9. Details of experimental methods, and additional results

For all inference methods, we used the open source libDAI library (Mooij, 2010) and averaged over 100 random models. We show results first for smaller models (complete graph on 10 variables and 5×5 grids), and then in Figure 11 for Bethe for larger models (complete graph on 15 variables and 9×9 grids). Wherever possible, we were consistent with the approaches of Weller and Domke (2016). We experienced difficulty with mean field (MF), since a randomly initialized run could return a very suboptimal solution. Hence, each time we used 100 random initializations and took the solution with highest estimate of the partition function (the most accurate since MF always provides a lower bound). Still, we experienced some convergence difficulties and advise caution in interpreting our MF results.

For MF,
`MF[tol=1e-7,maxiter=10000,damping=0.0,init=RANDOM,updates=NAIVE]`

For Bethe,
`HAK[doubleloop=1,clusters=BETHE,init=UNIFORM,tol=1e-7,maxiter=10000]`
This is guaranteed to converge to a stationary point of the Bethe free energy (whereas BP may not converge).

For TRW,
`TRWBP[updates=SEQFIX,tol=1e-7,maxiter=10000,logdomain=0,nrtrees=1000,...`
`damping=0.25,init=UNIFORM]`

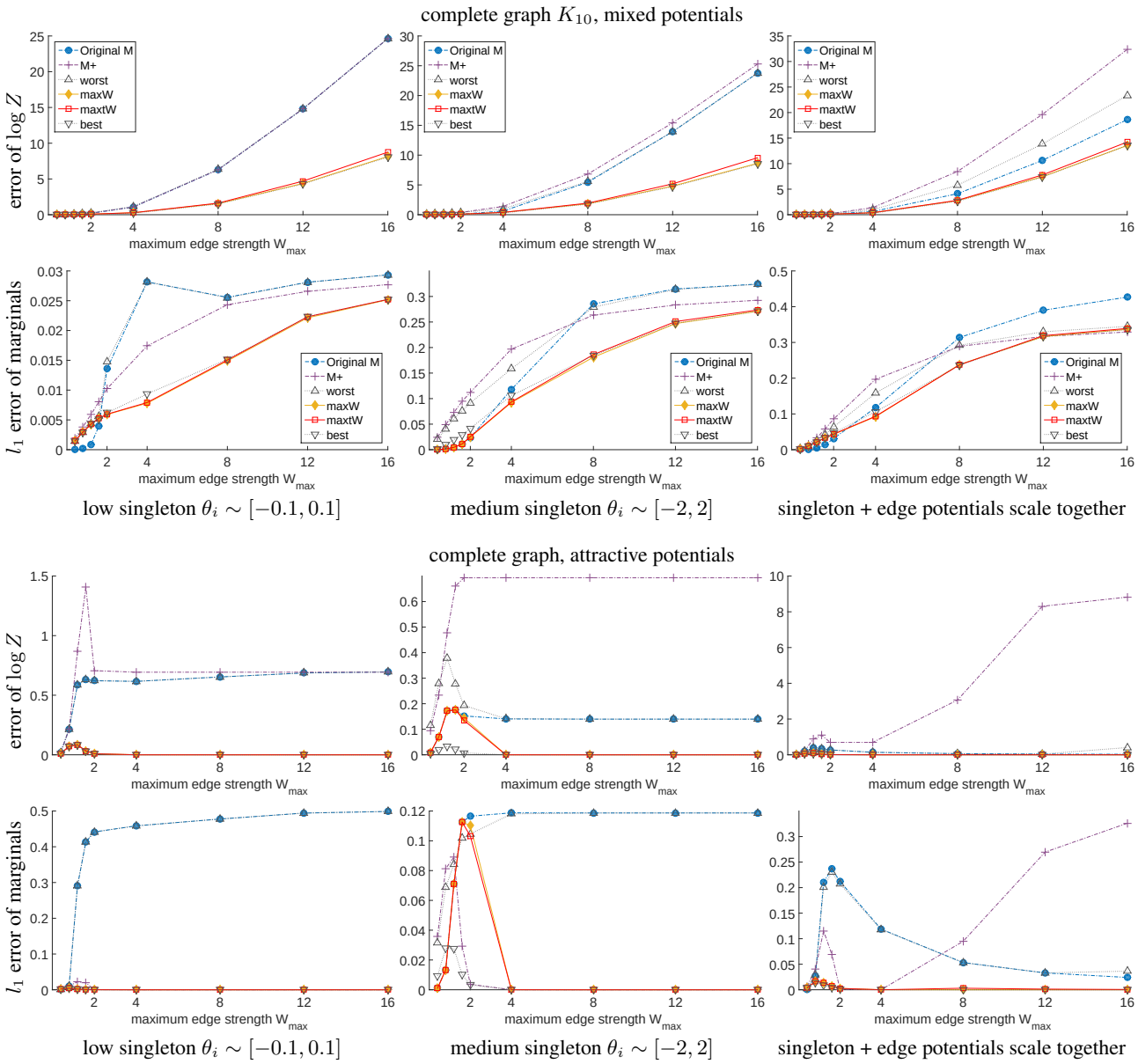

Figure 5. Average error plots over 100 runs for the Bethe approximation, complete graph with 10 variables

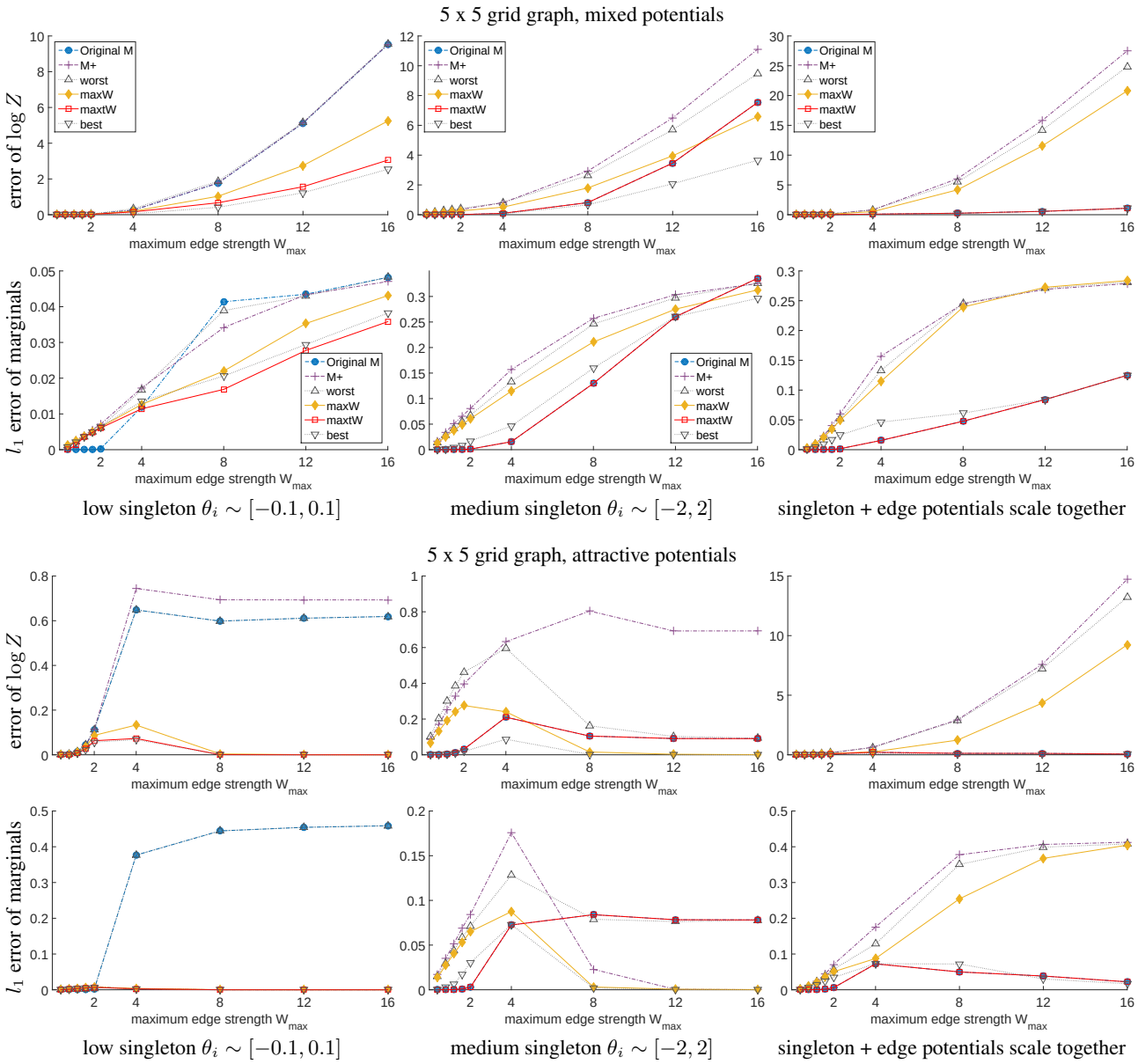

Figure 6. Average error plots over 100 runs for the Bethe approximation, 5 x 5 grid graph

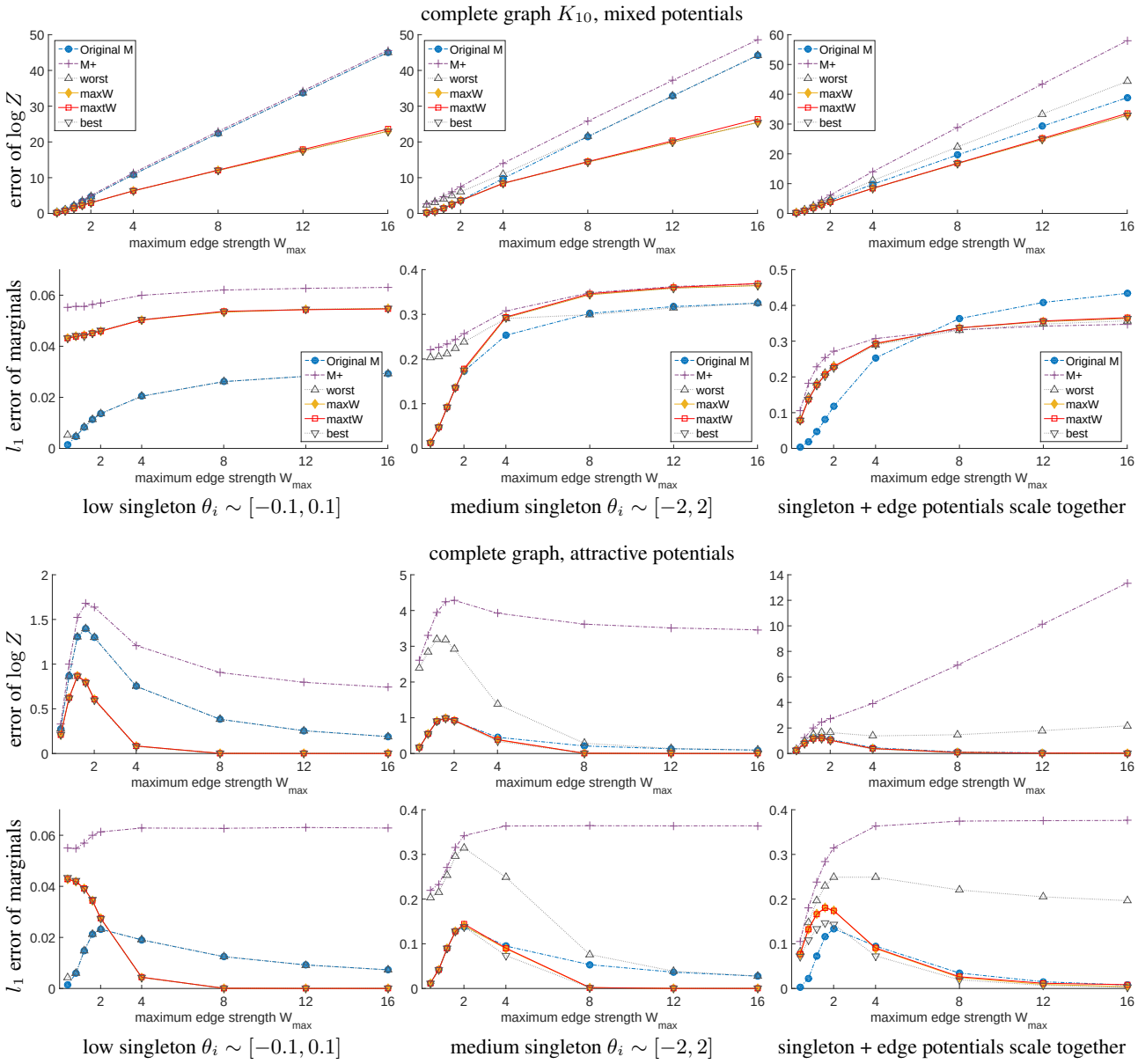

Figure 7. Average error plots over 100 runs for the TRW approximation, complete graph with 10 variables
Note the very low scale for l_1 error of marginals for low singleton potentials.

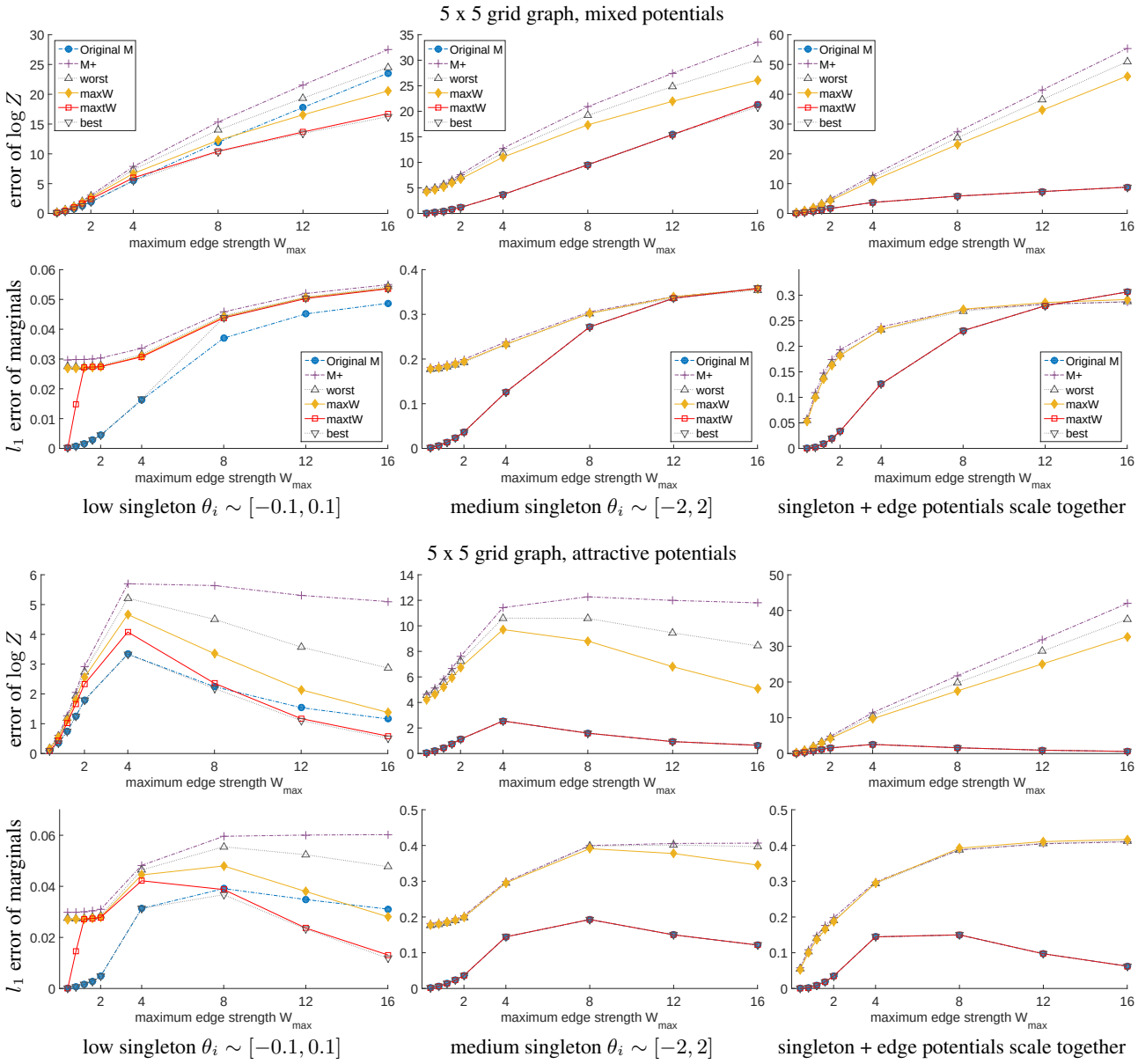

Figure 8. Average error plots over 100 runs for the TRW approximation, 5 x 5 grid graph
Note the very low scale for l_1 error of marginals for low singleton potentials.

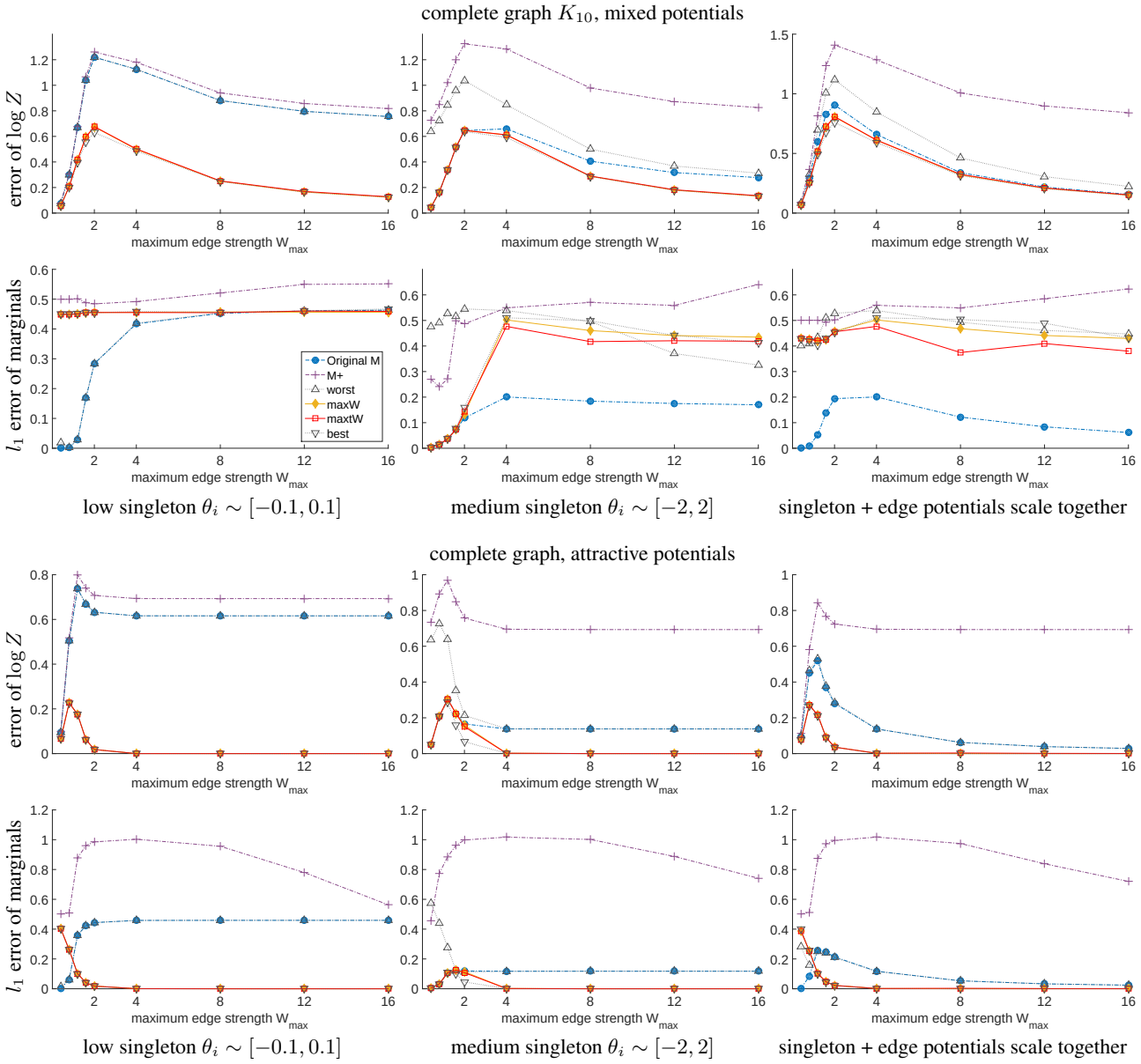

Figure 9. Average error plots over 100 runs for the MF approximation, complete graph with 10 variables

Results for the error of marginals for the complete graph look interesting and warrant further investigation, though we suspect these may be due to problems with our MF algorithm implementation.

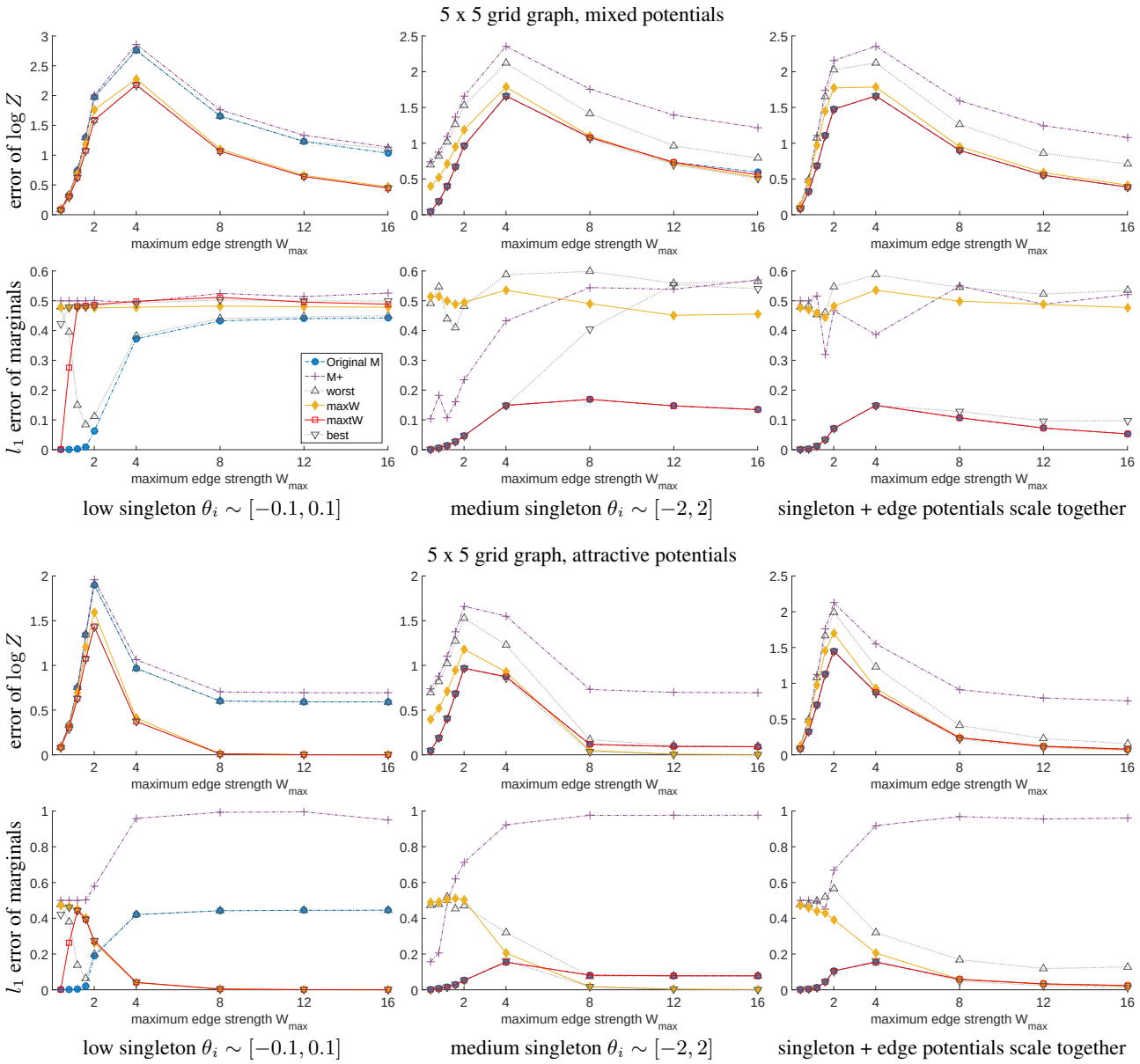

Figure 10. Average error plots over 100 runs for the MF approximation, 5 x 5 grid graph

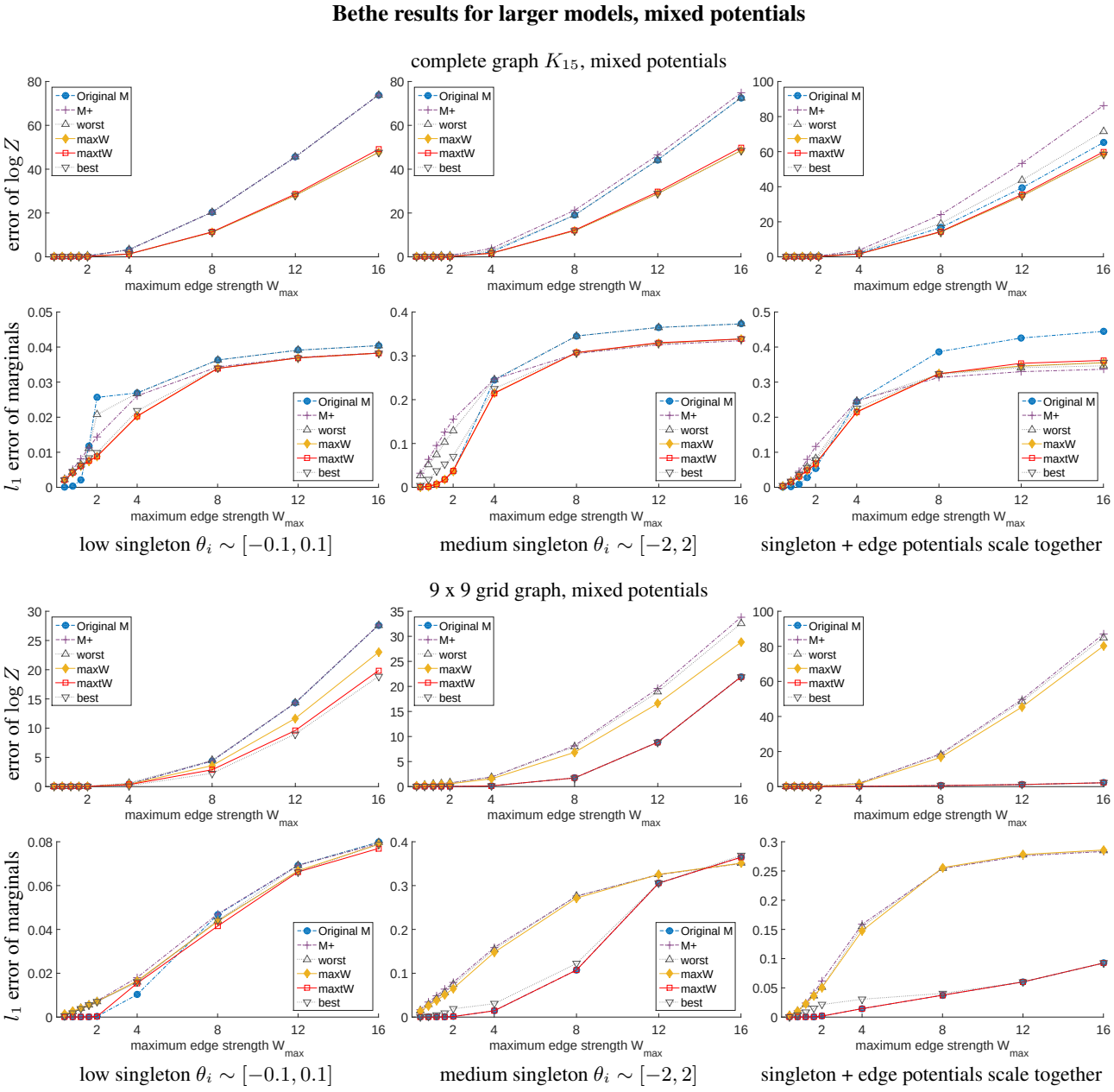

Figure 11. Average error plots over 100 runs for the Bethe approximation, mixed potentials