Characterizing Tightness of LP Relaxations by Forbidding Signed Minors

Adrian Weller
University of Cambridge
aw665@cam.ac.uk

Abstract

We consider binary pairwise graphical models and provide an exact characterization (necessary and sufficient conditions observing signs of potentials) of tightness for the LP relaxation on the triplet-consistent polytope of the MAP inference problem, by forbidding an odd-K_5 (complete graph on 5 variables with all edges repulsive) as a signed minor in the signed suspension graph. This captures signs of both singleton and edge potentials in a compact and efficiently testable condition, and improves significantly on earlier results. We provide other results on tightness of LP relaxations by forbidding minors, draw connections and suggest paths for future research.

1 INTRODUCTION

Discrete undirected graphical models play a central role in machine learning, providing a powerful and compact way to model relationships between variables. A key challenge is the combinatorial search problem to identify a most likely configuration of variables, termed *maximum a posteriori* (MAP) or *most probable explanation* (MPE) inference. This has received a great deal of attention from various communities, where it is sometimes framed as energy minimization (Kappes et al., 2013) or as solving a valued constraint satisfaction problem (VCSP, Schiex et al., 1995).

Since the problem is NP-hard, much work has attempted to identify restricted settings where polynomial-time methods are feasible. Where possible, we call such settings *tractable* and the methods *efficient*. Two types of restriction have been considered separately, either: (i) structural constraints on the topology of connections between variables; or (ii) families of potential functions.

Exploring the first theme, Chandrasekaran et al. (2008) showed that, if no restriction is placed on types of potentials, then the structural constraint of *bounded treewidth* is needed for tractable inference.[1] See §2-4 for all definitions.

Recent work (Kolmogorov et al., 2015; Thapper and Živný, 2015) has examined the power of using a linear programming (LP) relaxation of the discrete optimization problem. An LP attains an optimum at a vertex of the feasible region; if this vertex is integral, then it provides an exact solution to the original problem and we say that the LP is *tight*. If the LP is performed over the *marginal polytope*, which enforces global consistency (Wainwright and Jordan, 2008), then this LP is always tight, but exponentially many constraints are required, hence the method is not efficient. The marginal polytope is typically relaxed to the *local polytope* LOC, which enforces only pairwise consistency, requiring a number of constraints linear in the number of edges. Thapper and Živný (2015) showed that, if no restriction is placed on topology, then for a given family of potentials, either LP+LOC is tight, and hence solves all such problems efficiently, or the problem set is NP-hard.

Here we consider *hybrid* conditions (Cooper and Živný, 2011), which combine constraints on both structure and potentials, an exciting field with little prior work. Focusing on the important class of binary pairwise models,[2] and considering each edge to be *signed* as either *attractive* or *repulsive*, we establish precise hybrid characterizations for when certain LP relaxations will be tight for all valid potentials. By valid, we mean potentials that observe the signs (attractive or repulsive) of the edges. We show that these characterizations may be achieved by forbidding particular signed *minors* of the signed graph topology, yielding compact and efficiently testable conditions.

In applications, LP relaxations are widely used for structured prediction but the most common form, LP+LOC, often yields a fractional solution, motivating constraints for

[1] This result makes mild assumptions, specifically the grid-minor hypothesis (Robertson et al., 1994), and that NP $\not\subseteq$P/poly. See also (Kwisthout et al., 2010).

[2] Eaton and Ghahramani (2013) showed that any discrete graphical model may be either exactly represented, or arbitrarily well approximated, by a binary pairwise model, though the number of variables may increase substantially.

higher order cluster consistency (Batra et al., 2011). Weller et al. (2016) considered the LP relaxation over the *triplet-consistent* polytope TRI, which is the next level up from LOC in the hierarchy given by Sherali and Adams (1990) and is still efficient. Whereas it is known that LP+LOC is tight for any model which is *balanced*, Weller et al. (2016) showed that LP+TRI is tight for any model which is *almost balanced*. Further they demonstrated that almost balanced models may be 'pasted' together in certain configurations, while still guaranteeing tightness of LP+TRI.

The results of Weller et al. (2016) and our stronger characterizations here are very relevant to many problems in computer vision, such as foreground-background image segmentation, where due to contiguity of real objects, learned edges are mostly attractive, leading to a model which is 'close to balanced'. For example, on the horses dataset considered by Domke (2013), LP+LOC is loose but LP+TRI is often tight. Our work helps to explain this phenomenon.

We consider a refinement by examining the signs not only of edge potentials, but also of singleton potentials. These can be neatly incorporated by considering the signed *suspension graph* of a model, which adds one extra node with edges to the other variable nodes, each new edge corresponding to a singleton of the original model; see §4.

1.1 MAIN RESULTS

Our strongest result is Theorem 14, which shows that LP+TRI is tight for all valid potentials, observing signs of both edge and singleton potentials, iff the signed suspension graph does not contain an *odd-*K_5 (the complete graph on 5 nodes with all edges odd/repulsive, see §3) as a signed minor. This is a more powerful, signed version of an unsigned result, Theorem 13, that follows from the work of Barahona and Mahjoub (1986), showing that LP+TRI is tight for all valid unsigned potentials iff the unsigned suspension graph does not contain K_5 as an unsigned minor; see §5. For a sense of the additional power of the signed version, Theorem 14 allows models with arbitrarily high treewidth, provided only that there is no odd-K_5 minor (*one particular* signing of a K_5 structure), whereas Theorem 12 prohibits a K_5 minor of *any* type; see Table 2.

A weaker corollary of Theorem 14, our Theorem 10 shows that if we are less observant and do not examine singleton potentials, then LP+TRI is tight for all valid potentials (respecting signs of edges only) iff the signed graph topology does not contain an *odd-*K_4 as a signed minor.[3] This may be directly compared to the sufficient conditions of Weller et al. (2016), which similarly do not examine singleton potentials. We show that Theorem 10 is a significant improvement: it covers a substantially larger set of models, provides a compact condition that is both necessary and sufficient, and is efficiently testable; see §4.3.

As another consequence of Theorem 14, we obtain a result that may be of significant practical interest. Theorem 16 shows that in some cases, the number of cycle constraints needed to enforce integrality for LP+TRI may be dramatically reduced to just the *signed cycle constraints*; see §5.

We also reframe earlier results on tightness of LP relaxations in terms of forbidden minors. This perspective elegantly captures conditions on both structure and potentials, and reveals fascinating connections which prompt natural directions to explore in future work. See Table 2 for a summary and §6 for discussion.

1.2 APPROACH AND RELATED WORK

Characterizing properties by forbidden minors has been a fruitful theme in graph theory since the fundamental work of Robertson and Seymour, which builds over more than 20 papers to the *graph minor theorem*, described by Diestel (2010) as "among the deepest theorems that mathematics has to offer." We describe elements of the approach in §2, its extension to signed graphs, signed minors and *odd* minors in §3, and its relevance to LP relaxations in §4.

Odd-minor-free graphs have received attention in theoretical computer science, for example (Demaine et al., 2010). However, aside from the characterization by Watanabe (2011) of when belief propagation has a unique fixed point in terms of signed minors, to our knowledge there has been little direct use of this perspective in machine learning.[4]

To show our results, we connect several earlier themes, for which we provide relevant background. A key result by Guenin (2001) showed that a signed graph is *weakly bipartite*, which characterizes integrality of the vertices of a particular polyhedron, iff it does not contain an odd-K_5 minor; see §3. We draw on connections between: the marginal polytope of a model, with its LOC and TRI relaxations; and the corresponding *cut polytope* of its suspension graph, with its *rooted semimetric* RMET and *semimetric* MET relaxations (De Simone, 1990; Deza and Laurent, 1997); see §5.1. We use a link between MET and the *cycle inequalities* (Barahona, 1993); see §5.2. We extend these ideas to signed graph topologies in §5.3. Some of these connections for the unsigned case were considered by Sontag (2010).

2 GRAPHS AND MINORS

We follow standard definitions and omit some familiar terms. For more background, see (Diestel, 2010), particularly §12 for a survey of treewidth and forbidden minors.

[3] Theorem 14 allows any odd-K_4 minor in G provided it is not part of an odd-K_5 in ∇G, a much stronger result; see Table 2.

[4] Junction trees (Cowell et al., 1999), treewidth and unsigned graph minors are closely related. Treewidth was discussed by Halin but gained popularity through use by Robertson and Seymour (see historical note by Seymour, 2014).

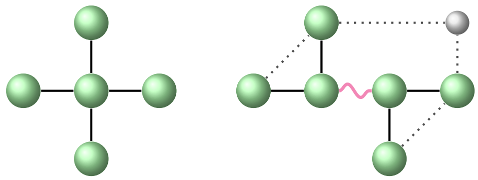


Figure 1: The left graph is a *minor* (unsigned) of the right graph, obtained by deleting the grey dotted edges and resulting isolated small grey vertex, and contracting the purple wavy edge. See §2.

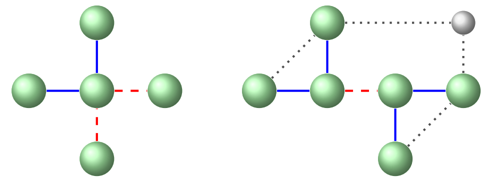


Figure 2: The left graph is a *signed minor* of the right signed graph, obtained similarly to Figure 1 except that before contracting the repulsive edge, first flip the vertex at its right end. Solid blue (dashed red) edges are attractive (repulsive). Grey dotted edges on the right are deleted and may be of any sign. See §3.2.

t	Forbidden minors for a graph to have treewidth $\leq t$
1	K_3
2	K_4
3	K_5 and 3 others
4	K_6 and more than 70 others

Table 1: Characterization of low values of bounded treewidth by forbidding minors.

A *graph* $G(V, E)$ is a set of vertices V, and undirected edges E, where each edge $(i, j) \in E$ connects i and $j \in V$. The *complete* graph on n vertices, written K_n, has all $\binom{n}{2}$ edges. A pairwise graphical model topology is always assumed to be a *simple* graph, that is a vertex may not be adjacent to itself (no loops) and each pair of vertices may have at most one edge (no multiple edges). However, when we consider *minors*, we allow loops and multiple edges.

A *minor* of a graph G is obtained from G by deleting edges or isolated vertices (as may be done to form a subgraph), or also by *contracting* edges. To contract edge (i, j) means to remove the vertices i and j, and replace them by a new vertex with edges to all remaining vertices that were previously adjacent to i or j. See Figure 1 for an example.

For any property $\mathcal{P}$ of a graph, we say that $\mathcal{P}$ is *closed under taking minors* (or *minor-closed*) if whenever G has property $\mathcal{P}$ and H is a minor of G, then H has $\mathcal{P}$.

A consequence of the *graph minor theorem* of Robertson and Seymour is the following deep result.

Theorem 1 (Robertson and Seymour, 2004)**.** *If a graph property $\mathcal{P}$ is closed under taking minors then it can be characterized by a* finite *set of forbidden minors, i.e. G has $\mathcal{P}$ iff G has none of the finite forbidden set as a minor.*

There are important examples of graph properties closed under taking minors where this finite set has just a few members. Perhaps the best known is the early result of Wagner (1937) that a graph G is planar iff G does not contain K_5 or $K_{3,3}$ as a minor ($K_{3,3}$ is the complete bipartite graph where each partition has 3 vertices).

Another property closed under taking minors is bounded *treewidth*. A definition of treewidth of a graph G that may be familiar from the junction tree construction is that it is one less than the minimum possible size of a largest clique in a triangulation of G (Wainwright and Jordan, 2008). The forbidden minors are known for low values of bounded treewidth, see Table 1. For example, a tree has treewidth 1 and cannot contain a K_3 minor.

Robertson and Seymour also showed that checking for any fixed minor may be performed efficiently.

Theorem 2 (Robertson and Seymour, 1995)**.** *For any fixed graph H and a given graph G with n vertices, there is an $O(n^3)$-time algorithm to determine if H is a minor of G.*[5]

Together, Theorems 1 and 2 show that any minor-closed graph property may be decided in polynomial-time.

3 SIGNED GRAPHS & SIGNED MINORS

A *signed graph* (Harary, 1953) is a graph (V, E) together with one of two possible signs for each edge. This is a natural structure when considering binary pairwise models, where we characterize edges as either *attractive* (or *even*) or *repulsive* (or *odd*), see §4. Where helpful for clarity, we refer to the standard graphs of §2 as *unsigned* graphs. We shall see that important concepts and results for minors of unsigned graphs have corresponding results for signed minors of signed graphs.

In a signed graph, a fundamental property of any cycle is whether or not it is a *frustrated cycle* (or *odd cycle*), i.e. if it is a cycle with an odd number of repulsive (or odd) edges. A signed graph is *balanced* if it contains no frustrated cycles. Following Weller (2015), a signed graph is *almost balanced* if it contains a vertex such that deleting it renders the remaining graph balanced.

3.1 FLIPPING/RESIGNING AND ODD GRAPHS

Given a signed graph, a subset of variables $S \subseteq V$ may be *flipped* (or *switched*). This flips the sign of any edge with exactly one end in S (i.e. flips the edge between attractive/even and repulsive/odd), and is called a *resigning*. It is easily seen that this operation does not change the nature (frustrated or not) of any cycle. For binary graphical models, this has a natural interpretation: if the original model has variables $\{X_i \in \{0, 1\} : i \in V\}$ then consider an

[5] Kawarabayashi et al. (2012) improved this to $O(n^2)$-time.

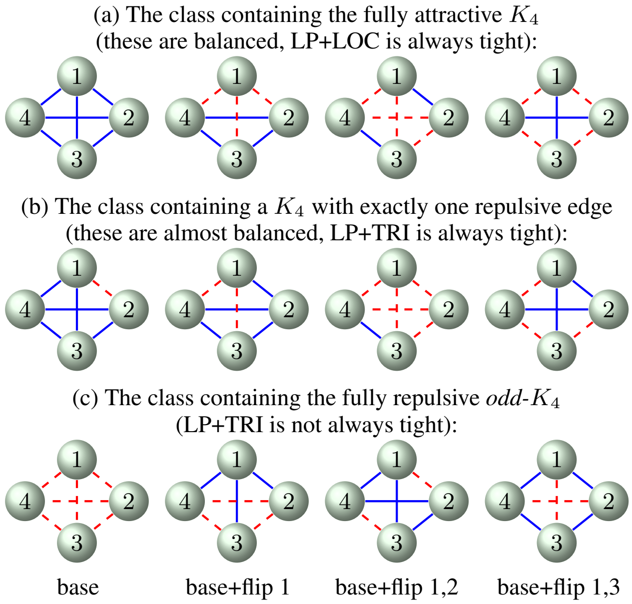


Figure 3: Examples of signed K_4 graphs. These are complete graphs on 4 vertices where each edge is either attractive/even (solid blue) or repulsive/odd (dashed red). Each row illustrates examples from one of the three signing equivalence classes. At bottom left is an *odd*-K_4. See §3.

equivalent model with variables $\{Y_i : i \in V\}$ given by $Y_i = 1 - X_i$ for $i \in S$, $Y_i = X_i$ for $i \in V \setminus S$, with new potentials set to match properties of the original model.[6]

Two signed graphs are *signing equivalent* (sometimes called *gauge equivalent*) if they are isomorphic up to a resigning (an equivalence relation). We are interested in signed graphs only up to signing equivalence; see §3.2.

For any unsigned graph G, the signed graph *odd*-G is the signed version of G where every edge is odd (or repulsive). Figure 3 shows examples of signed K_4 graphs, in their signing equivalence classes. The bottom row shows an *odd*-K_4 at the left, together with possible resignings.

3.2 SIGNED MINORS

A *signed minor* of a signed graph is obtained just as for an unsigned minor of an unsigned graph with the following modifications: any resigning operations are permitted (see §3.1); and contractions are allowed only for attractive/even edges. Note that a repulsive/odd edge may first be resigned to an attractive/even edge by flipping either end vertex (which will also affect its other incident edges) and then contracted. See Figure 2 for an example, which may be compared to the unsigned minor example of Figure 1.

A significant project is in progress to try to generalize all of Robertson and Seymour's graph minor theory to the much broader class of Γ-labeled graphs for any finite abelian group Γ (Geelen et al., 2014), which includes signed graphs by considering $\Gamma = \mathbb{Z}/2\mathbb{Z}$. An equivalent result to Theorem 1 is claimed, though the formal write-up is still to come. An equivalent result to Theorem 2 has been shown.

Theorem 3 (A special case of Theorem 1.1.10 of Huynh, 2009)**.** *For any fixed signed graph H, there is a polynomial-time algorithm which determines if an input signed graph G contains H as a signed minor.*

3.3 WEAKLY BIPARTITE SIGNED GRAPHS

If a signed graph is balanced, its vertices may be partitioned into two exhaustive groups s.t. all inter-group edges are odd and all intra-group edges are even (Harary, 1953); the resigning obtained by flipping either group renders all edges even. With this observation, a signed graph which is balanced is sometimes called *bipartite* (related, but different, to the standard meaning of bipartite for unsigned graphs).

Generalizing bipartite signed graphs, a signed graph $G(V, E)$ with edge signs is *weakly bipartite* if the following polyhedron Q has only integral vertices:

$$Q = \left\{ y \in \mathbb{R}_+^{|E|} : \sum_{e \in D} y_e \geq 1, \forall \text{ odd cycles } D \text{ of signed } G \right\} \tag{1}$$

Here, odd cycles D are in the signed sense, i.e. have an odd number of odd edges. We shall see in §5.3 that Q relates closely to the triplet-consistent polytope TRI of a graphical model, if we consider signs of all potentials. We make use of the following result, which proved a conjecture of Seymour (1977), earning Guenin a Fulkerson prize in 2003.

Theorem 4 (Guenin, 2001)**.** *A signed graph is weakly bipartite iff it does not contain an odd-K_5 as a signed minor.*[7]

4 GRAPHICAL MODELS AND LP RELAXATIONS

We consider a binary pairwise undirected model with n variables $X_1, \ldots, X_n \in \{0, 1\}$. Let $x = (x_1, \ldots, x_n) \in \{0, 1\}^n$ be one complete configuration. The probability distribution is specified by $p(x) \propto \exp[\text{score}(x)]$, where we choose a symmetric minimal reparameterization (Wainwright and Jordan, 2008) such that

$$\text{score}(x) = -\sum_{i \in V} \theta_i x_i - \sum_{(i,j) \in E} W_{ij} \mathbb{1}[x_i \neq x_j], \tag{2}$$

where $\mathbb{1}[\cdot]$ is the indicator function. The model's unsigned topology is the graph $G(V, E)$, with n variables $V = \{1, \ldots, n\}$ and $m = |E| \leq \binom{n}{2}$ edge relationships between the variables. The n singleton parameters $\{\theta_i : i \in V\}$ and m edge weights $\{W_{ij} : (i, j) \in E\}$

[6] Given the form of potentials (2) we choose in §4, this flips the signs of $\{\theta_i : i \in S\}$ and $\{W_{ij} :$ exactly one of i or $j \in S\}$.

[7] The original proof is long. A shorter proof was provided by Schrijver (2002). Both proofs rely on a result of Lehman (1990).

define the potentials, which we allow to take any rational value (to enable polynomial-time algorithms).

In addition to the unsigned graph G, we shall be interested in two more informative ways of considering a model's topology. The *signed graph* G assigns edge signs according to the signs of edge potentials. If $W_{ij} > 0$, the edge (i, j) tends to pull X_i and X_j toward the same value and is *attractive* (or *even*). If $W_{ij} < 0$, the edge is *repulsive* (or *odd*).

The *signed suspension graph* $\nabla G(V', E')$ of a model adds an extra node 0, that is $V' = V \cup \{0\}$. Edges to 0 encode singletons of the model, with $E' = E \cup \{(0, i) : \theta_i \neq 0\}$.

With ∇G in mind, we have chosen the form of (2) carefully, using negative signs so $W_{ij} > 0$ is attractive, and $\mathbb{1}[x_i \neq x_j]$ edge terms in order to facilitate later demonstration of the equivalence between the MAP problem and a max cut problem on the edge-weighted suspension graph; see §5.

In ∇G, it may be helpful to consider the added node 0 as being set to the value 0; then regarding (2), the singleton potential terms $\theta_i x_i$ may be viewed as $\theta_i \mathbb{1}[x_i \neq 0]$, and hence all singleton and edge potential terms follow the same sign convention. In particular, the sign of each new edge $(0, i)$ in ∇G matches that of θ_i: if $\theta_i > 0$ then the added edge is attractive, pulling X_i toward 0; if $\theta_i < 0$ then $(0, i)$ is repulsive (or odd).

4.1 LINEAR PROGRAMMING FOR MAP

The potential parameters may be concatenated to form a vector $w = (-\theta_1, \dots, -\theta_n, \dots, -W_{ij}, \dots) \in \mathbb{Q}^d$, where $d = n + m$. Let $y_{ij} = \mathbb{1}[x_i \neq x_j]$, and for any configuration x, similarly concatenate the n x_i and m $y_{ij}(x)$ terms into a vector $z = (x_1, \dots, x_n, \dots, y_{ij}, \dots) \in \{0, 1\}^d$. Now $\text{score}(x) = w \cdot z$, yielding the following integer linear programming formulation for MAP inference, to identify

$$z^* \in \operatorname*{arg\,max}_{z: x \in \{0,1\}^n} w \cdot z \tag{3}$$

The convex hull of the 2^n possible integer solutions in $[0, 1]^d$ is the *marginal polytope* $\mathbb{M}$ for our choice of singleton and edge terms.[8] Regarding the convex coefficients as a probability distribution p over all possible states, $\mathbb{M}$ may be considered the space of all singleton and pairwise mean marginals that are consistent with some global distribution p over the 2^n states, that is

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

A standard approach is to relax (3) to a linear program (LP). Performing this over $\mathbb{M}$ remains intractable since the number of linear constraints required grows extremely rapidly with n (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 (3), 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 $\mathcal{L}_n = \mathbb{M} \subseteq \mathcal{L}_{n-1} \subseteq \cdots \subseteq \mathcal{L}_3 = \text{TRI} \subseteq \mathcal{L}_2 = \text{LOC}$. Clearly LP+$\mathcal{L}_n$ is always tight. The following result, derived using the junction tree theorem (Cowell et al., 1999), gives a sufficient condition for tightness at any order.

Theorem 5 (Wainwright and Jordan, 2004)**.** *If the graph of a model has treewidth $\leq r - 1$ then LP+$\mathcal{L}_r$ is tight.*

4.2 RELATION TO MINORS, NEW RESULTS

Theorem 5 provides a sufficient condition that considers only the treewidth of the unsigned graph G, without any regard to the potentials. As remarked in §2, the graph property of bounded treewidth is minor-closed, hence can be characterized by excluding a finite set of forbidden minors, see Table 1 for examples.

We now make the following observation, where "valid potentials" for a graph means any potentials that respect the graph structure (signed or unsigned accordingly).

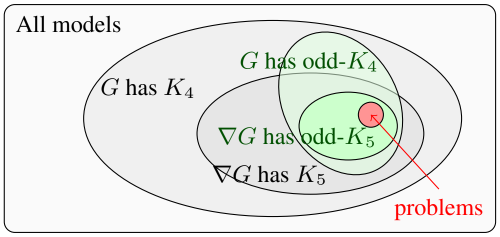
Theorem 6. *The property $\mathcal{P}$ of a graph G that "LP+$\mathcal{L}_r$ is tight for all valid potentials on G" is minor-closed, whether G is unsigned or signed (if signed then use signed minors).*

Proof. The property $\mathcal{P}$ is maintained under deletion, contraction and (for signed graphs) resigning. To see this for contraction: if an edge (i, j) of G is contracted to yield G', then for any valid model M' on G', consider the model M on G which has all the same potentials and in addition set the edge potential for (i, j) to be sufficiently high such that in M this forces X_i and X_j to take the same value. □

Hence, we should expect to be able to characterize LP tightness for all valid potentials, for both unsigned and signed topologies, by specifying a finite set of forbidden minors (signed minors in the signed case), see §2 and §3.

From Theorem 5 and Table 1: if we consider only the unsigned topology G, then LP+LOC (LP+TRI) is tight for all potentials if the graph G does not contain a K_3 (K_4, respectively) as a minor. To demonstrate the converse, and as a result of independent interest, we show the following.

[8] Our choice of edge term $\mathbb{1}[x_i \neq x_j]$ will facilitate later analysis of ∇G in §5. A common alternative choice for edges is to use $x_i x_j$, which leads to an equivalent polytope, sometimes called the *Boolean quadric polytope* QP^n (Padberg, 1989).



		Forbidden minors			
Sherali-Adams cluster size		Graph G		Suspension graph ∇G	
		Unsigned	Signed	Unsigned	Signed
LOC	$\mathcal{L}_2$	K_3	odd-K_3		
TRI Theorem	$\mathcal{L}_3$	K_4 Thm 8	odd-K_4 Thm 10	K_5 Thm 13	odd-K_5 Thm 14
	$\mathcal{L}_4$	K_5+?	odd-K_5+?	K_6+?	odd-K_6+?

Table 2: Summary of results characterizing tightness of LP relaxations by forbidden minors. All conditions may be checked efficiently. **Right:** The section for TRI (shaded blue) contains our main new results: Theorem 10 for signed G; and the stronger Theorem 14 for signed ∇G, which examines singleton and edge potentials. Theorem 14 implies Theorems 13, 10 and 8. Results for $\mathcal{L}_4$ are unknown. **Left:** Illustration of the model classes for LP+TRI, where *problems* are models for which LP+TRI is not tight. Theorem 14 is the most powerful result, showing that all problems lie within the set of models where ∇G contains an odd-K_5. See discussion in §4.3 and §6.

Theorem 7. *LP+$\mathcal{L}_r$ is not tight for the fully connected model on $n = r+1$ variables with all $\theta_i = W_{ij} = -1\ \forall i \in V, (i,j) \in E$. Note that this model has signed G which is an odd-K_{r+1}, and signed ∇G which is an odd-K_{r+2}.*

Proof. The proof is from first principles. Consider the distribution for each r-cluster that is uniform over all configurations with $\lfloor \frac{r}{2} \rfloor$ 0s and $\lceil \frac{r}{2} \rceil$ 1s. This has higher score than the best integral configuration of $\lfloor \frac{n}{2} \rfloor$ 0s and $\lceil \frac{n}{2} \rceil$ 1s. □

Applying Theorem 7 for $r = 2$ and 3 yields the following result (since if a model contains a K_{r+1} minor, then we may assume potentials such that the model is the K_{r+1}).

Theorem 8. *Considering unsigned topologies: LP+LOC is tight for all valid potentials iff G does not contain a K_3 minor; LP+TRI is tight for all valid potentials iff G does not contain a K_4 minor.*

We next provide stronger results by considering the signs of edge potentials. Intriguingly, both for LOC and TRI, the forbidden signed minor is exactly the odd version of the forbidden unsigned minor in Theorem 8.

Theorem 9. *LP+LOC is tight for all valid potentials iff the signed graph G does not contain an odd-K_3 signed minor.*

Proof. It is easily seen that if the signed graph of a model does not contain an odd-K_3 signed minor, then it is balanced (see §3). Earlier work showed that LP+LOC is tight for any balanced model (Padberg, 1989; Weller et al., 2016). Necessity follows from Theorem 7. □

Theorem 10. *LP+TRI is tight for all valid potentials iff the signed graph G does not contain an odd-K_4 signed minor.*

Theorem 10 follows as a corollary of the stronger Theorem 14, which we shall show in §5, which also considers the signs of singleton potentials by examining the signed suspension graph ∇G (defined in §4 just before §4.1).

4.3 REMARKS, EARLIER WORK

Taken together, our results in §4.2 employ the framework of forbidden minors to characterize compactly the tightness of LP relaxations in an elegant and unifying way. See Table 2 for a summary, which includes later results from §5. Note the interesting relationships across conditions, all of which may be checked efficiently by Theorems 2 and 3.

Little was known theoretically about conditions for tightness of LP+TRI before Weller et al. (2016) showed that it was sufficient for a model to be almost balanced (defined in §3). They also demonstrated a composition result which allows almost balanced sub-models to be pasted together in particular ways, while maintaining tightness. The condition in Theorem 10 is substantially stronger: the new condition is both necessary and sufficient; it is compact to describe and it is efficient to check. Further, we next show that Theorem 10 covers a strict superset of models.

Observe that the property for a signed graph of being almost balanced is closed under taking signed minors. An odd-K_4 is clearly not almost balanced, see Figure 3. Hence almost balanced models $\subseteq$ models whose signed topology does not contain an odd-K_4 minor. Next consider the permitted pasting operation (Weller et al., 2016, Theorem 12), which allows sub-models to be pasted together either on a single variable or, in limited settings, on an edge. If each sub-model is odd-K_4-free, then so too is the pasted combination. Hence, Theorem 10 covers all the models covered by Weller et al. (2016); we next show that Theorem 10 covers a significant additional class of models.

Signed graphs that do not contain an odd-K_4 minor have been studied previously (Gerards, 1988; Truemper, 1998). An important class that is odd-K_4-minor-free but not almost balanced is planar signed graphs with two odd faces (i.e. all but exactly two faces have even bounding cycles). See Figure 4 for an example by J. Carmesin which is 3-connected, hence there is no way it could be constructed by pasting almost balanced sub-models on edges or vertices.

5 INCLUDING SINGLETONS, ∇G

In this Section, we extend the analysis of §4 to include singleton potentials, by now considering the suspension graph ∇G rather than just the base graph G. We build to §5.3,

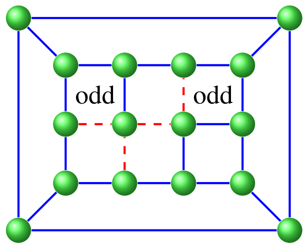


Figure 4: A 3-connected signed graph that is not almost balanced, hence the results of Weller et al. (2016) cannot be used to show that LP+TRI is tight for such a model; yet it is odd-K_4-minor-free, hence Theorem 10 proves that LP+TRI is tight for all valid potentials. This is a planar signed graph with two odd faces (Gerards, 1988, §3.2), with the odd faces indicated (others are even); see §4.3. Solid blue (dashed red) edges are even (odd).

Potential values for n variables + m edges	$\leftrightarrow$	Potential values for $n+m$ edges
Marginal polytope $\mathbb{M}$	$\leftrightarrow$	CUT polytope of ∇G
TRI relaxation	$\leftrightarrow$	MET relaxation = CYC
LOC relaxation	$\leftrightarrow$	RMET relaxation

Table 3: Relations between polytopes; see §5.

where we state and prove our strongest result, Theorem 14, which characterizes tightness of LP+TRI if we examine the signs of both edge and singleton potentials. We show that this result implies Theorem 10 from §4, which examines only edge signs. Our approach relies on Theorem 4 (Guenin's result), connecting to it by showing relations between various polytopes.

In §4, we introduced the marginal polytope $\mathbb{M}$, together with its relaxations TRI and LOC, with $\mathbb{M} \subseteq$ TRI $\subseteq$ LOC. Here, we first show equivalences to the *cut polytope* CUT of the suspension graph ∇G, together with its relaxations MET (the *semimetric polytope*) and RMET (the *rooted semimetric polytope*) with CUT $\subseteq$ MET $\subseteq$ RMET, see Table 3. In §5.2, we relate MET to the *cycle inequalities* (Barahona and Mahjoub, 1986; Barahona, 1993) and provide Theorem 13, which does not consider signs of potentials. For more background, see (Deza and Laurent, 1997).

In §5.3 we consider signed cycles, then by combining results, we prove Theorem 14, our strongest new result. In addition, we are able to show that many typically used cycle inequalities may be redundant for enforcing integrality.

5.1 MARGINAL AND CUT POLYTOPES, AND THEIR RELAXATIONS

Here we establish many of the equivalences of polytopes shown in Table 3. Recall the definition of the suspension graph $\nabla G(V', E')$ from §4.

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

The *cut polytope* (Barahona, 1983; Barahona and Mahjoub, 1986) of ∇G is the convex hull of all such cut vectors, that is CUT = conv $\{\delta(S) : S \subseteq V'\}$. Although there are 2^{n+1} choices of S, CUT has 2^n vertices since by definition $\delta(S) = \delta(V' \backslash S)$. In fact, there is a linear bijection between CUT and $\mathbb{M}$, which is particularly simple given the form we selected for edge marginals in (4).[9] Given $d \in$ CUT with entries d_{ij} for $(i,j) \in E'$, d maps to $\mu \in \mathbb{M}$ where $\mu_j = d_{0j}$ for $j \in V$, and $\mu_{ij} = d_{ij}$ for $(i,j) \in E$.

To see this, d_{ij} may be interpreted as the marginal probability that $i, j \in V'$ lie in different partitions. Similarly, $\mu_{ij} \in \mathbb{M}$ is the marginal probability that X_i and X_j take different values; and μ_i is the probability that $X_i \neq 0$ (corresponding in ∇G to i being in a different partition to 0).

MAP inference for the model on G is now clearly equivalent to the max cut problem for ∇G, i.e.[10]

$$\max_{\mu \in \mathbb{M}} w \cdot \mu = \max_{d \in \text{CUT}} w' \cdot d, \;\; w'_{ij} = \begin{cases} -\theta_j & i = 0 \\ -W_{ij} & (i,j) \in E. \end{cases} \quad (5)$$

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, as shown in Table 3. 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):

$$\text{MET} \quad \forall i,j,k \in V', \;\; d(i,j) - d(i,k) - d(j,k) \leq 0 \quad (6)$$
$$d(i,j) + d(i,k) + d(j,k) \leq 2.$$

Remarkably, the constraints for RMET, the *rooted* triangle inequalities, are exactly just those of (6) for which one of i, j, k is 0, the vertex that was added 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* (Weller, 2016). See discussion in §6.

5.2 CYCLE INEQUALITIES, CYC POLYTOPE

Here we define the cycle inequalities and provide background showing how they may be used to characterize tightness of LP+TRI by forbidding unsigned K_5 as a minor of the unsigned suspension graph $\nabla G(V', E')$.

For any edge set $F \subseteq E'$ and $x \in [0,1]^{|E'|}$, let $x(F) = \sum_{e \in F} x(e)$. Let $C \subseteq E'$ be the edge set of a cycle in ∇G.

[9] If instead, edge terms of the form $x_i x_j$ are used for the marginal polytope, as in the Boolean quadric polytope (see footnote 8), then the linear bijection required is slightly more complex and is called the *covariance mapping* (De Simone, 1990).

[10] The negative signs before θ_j and W_{ij} terms are because we followed the convention that $W_{ij} > 0$ is an attractive edge, and made the signs of singleton potentials consistent; see §4.

At any vertex of CUT, if we traverse C, we must change partitions an even number of times.

Hence, the following *cycle inequalities* hold $\forall x \in$ CUT: for any cycle C and any edge subset $F \subseteq C$ with $|F|$ odd,

$$x(F) - x(C \setminus F) \leq |F| - 1. \tag{7}$$

Let CYC be the polytope defined by these constraints, i.e. CYC $= \{x \in [0,1]^{|E'|} : x(F) - x(C \setminus F) \leq |F| - 1$ for any cycle C of ∇G and any $F \subseteq C, |F|$ odd$\}$. The triangle inequalities (6) are special cases of (7) with $|C| = 3$, though note that those apply in MET to any triplet $i, j, k \in V'$ without regard to the edges E', whereas cycle inequalities apply only to the cycles of $\nabla G(V', E')$. Nevertheless, in fact, the following result holds; see Table 3.

Theorem 11 (Barahona and Mahjoub, 1986; Barahona, 1993)**.** *MET = CYC.*

Barahona and Mahjoub (1986) established the following important characterization for when the cycle inequalities are sufficient for tightness by forbidding a K_5 minor.

Theorem 12 (Barahona and Mahjoub, 1986)**.** *CUT = CYC iff unsigned ∇G does not contain K_5 as a minor.*

Using the equivalences of §5.1 (see (5) and Table 3), Theorems 11 and 12 together show the following result, which characterizes when LP+TRI is tight if we examine only the *unsigned* suspension graph ∇G.

Theorem 13. *LP+TRI is tight for all valid potentials iff unsigned ∇G does not contain K_5 as a minor.*

Theorems 11 and 12 of Barahona and Mahjoub (1986) are often used to show only that LP+TRI is tight for a planar model with no singleton potentials (which excludes both K_5 and $K_{3,3}$, see §2), e.g. see Theorem 3.3.2 in (Sontag, 2010). However, Theorem 13 is stronger, and perhaps is more naturally viewed instead as extending the characterization of treewidth ≤ 2 as K_4-minor-free; see §6.

5.3 SIGNED CYCLES, MISS POLYTOPE

Here we shall prove Theorem 14, a stronger, signed version of Theorem 13. Theorem 10 will follow as a corollary. For cycles in ∇G, to avoid confusion, we write C for the edge set of an unsigned cycle, and D for the signed edge set of a signed odd cycle (which has an odd number of odd edges).

Given results in §5.1-5.2 (see (5) and Table 3), we have

$$\max_{\mu \in \text{TRI}} w \cdot \mu = \max_{x \in \text{CYC}} w' \cdot x, w'_{ij} = \begin{cases} -\theta_j & i = 0 \\ -W_{ij} & (i,j) \in E. \end{cases} \tag{8}$$

We shall relate this to Theorem 4 (Guenin's result on weakly bipartite graphs from §3.3) to prove the following.

Theorem 14. *LP+TRI is tight for all valid potentials, observing signs of both edge and singleton potentials, iff the signed suspension graph ∇G does not contain odd-K_5 as a signed minor.*

Proof. We first show sufficiency of the condition. Regarding (8), CYC is defined by the inequalities (7), which we rewrite as $|F| - x(F) + x(C \setminus F) \geq 1$, or

$$\sum_{e \in F} (1 - x_e) + \sum_{e \in C \setminus F} x_e \geq 1. \tag{9}$$

The unsigned cycle inequality (9) applies for every cycle C of ∇G and every $F \subseteq C$ with $|F|$ odd. Aiming to relate (9) to the definition of a weakly bipartite graph (1), we introduce the following MISS polytope, which is equivalent to CUT by a reflection that adjusts for the signs of edges of $\nabla G(V', E')$. Recall how these signs are set in §4, and regarding (8), observe that edge $e \in E'$ is odd iff $w'_e > 0$.

Given a configuration of variables $X_1, \ldots, X_n \in \{0,1\}^n$, the corresponding vertex $\in \{0,1\}^{|E'|}$ of the CUT polytope has a 1 for edge $(i,j) \in E'$ iff $X_i \neq X_j$, taking $X_0 = 0$.

For MISS, instead the corresponding vertex $m \in \{0,1\}^{|E'|}$ has a 1 for edge $(i,j) \in E'$ iff m_{ij} 'misses' the potential score benefit that the edge offers. That is, take $X_0 = 0$ as before, and now for all $(i,j) \in E'$, assign $m_{ij} = 1$ if $X_i \neq X_j$ and the edge is even (attractive), or if $X_i = X_j$ and the edge is odd (repulsive); otherwise $m_{ij} = 0$.

Each of CUT and MISS is formed as the convex hull of its 2^n vertices. MISS is the reflection of CUT across $\frac{1}{2}$ in exactly the dimensions for which edges of ∇G are odd. That is, $x \in$ CUT maps bijectively to $y \in$ MISS, where $y_e = x_e$ for even edges, and $y_e = 1 - x_e$ for odd edges.

Let D be the edge set of an odd cycle of signed ∇G (i.e. D has an odd number of odd edges). Given any configuration of variables, as we go round D, to return to the same value, we must 'miss' at least one edge. That is, for any vertex $m \in$ MISS, $\sum_{e \in D} m_e \geq 1$. Hence, what we call the *signed cycle inequalities* hold $\forall y \in$ MISS $\subseteq [0,1]^{|E'|}$:

$$\sum_{e \in D} y_e \geq 1, \quad \forall \text{ odd cycles } D \text{ of signed } \nabla G. \tag{10}$$

Note the direct correspondence between the signed cycle inequalities (10) and the inequalities defining the weakly bipartite polyhedron Q (1). Observe the following.

Lemma 15. *Each signed cycle inequality* (10) *corresponds to an unsigned cycle inequality* (9)*.*

Proof. Given a signed cycle inequality (10), let F be the odd edges of D, with $|F|$ odd. Let $C = D$. The equivalent reflected form of (10) for $x \in$ CUT is $\sum_{e \in F} (1 - x_e) + \sum_{e \in C \setminus F} x_e \geq 1$, which matches (9) as required. □

Let CYC^R be the polytope CYC reflected in the odd edge dimensions, just as MISS may be obtained from CUT, so

MISS $\subseteq$ CYC^R as CUT $\subseteq$ CYC. Consider (8), note edge e is odd $\Leftrightarrow w'_e > 0$, let $x \in$ CYC map to $y \in$ CYC^R, then

$$\max_{\mu \in \text{TRI}} w \cdot \mu = \max_{y \in \text{CYC}^R} \sum_{\text{odd } e \in E'} w'_e(1-y_e) + \sum_{\text{even } e \in E'} w'_e y_e$$
$$= A + \max_{y \in \text{CYC}^R} w'' \cdot y, \quad (11)$$

where $A = \sum_{e: w'_e > 0} w'_e$ is a constant, and $w''_e = -|w'_e| \; \forall e \in E'$.

We are now ready to apply Theorem 4. We have {all valid integer solutions} $\subseteq$ MISS $\subseteq$ CYC^R, while (by Lemma 15) CYC^R enforces all the signed cycle inequalities (10), which match the weakly bipartite conditions (1). Further, in (11) we are maximizing an objective with every coefficient negative, which is needed since Q is a polyhedron unbounded in the positive directions (1). Finally, no invalid integer solutions lie in CYC^R. Hence, by Theorem 4, if signed ∇G does not contain odd-K_5 as a signed minor, then LP+TRI is tight.

For necessity of the condition, if ∇G does contain an odd-K_5 minor, then by choice of potentials, we may assume signed $\nabla G(V', E')$ to be exactly odd-K_5. Then Theorem 7 with $r = 3$ provides an example where LP+TRI is not tight. This completes the proof of Theorem 14. □

Corollaries. Theorem 14 may be used to prove Theorem 10 as follows. First, if signed G does not contain an odd-K_4, then clearly signed ∇G cannot contain an odd-K_5, hence LP+TRI is tight by Theorem 14. Now, if signed G does contain an odd-K_4, then we may select all negative singleton potentials for those variables, then use the example above for odd-K_5 in signed ∇G.

In practice, LP+TRI is often implemented by enforcing the (unsigned) cycle constraints (9) rather than all triplet constraints (Sontag, 2010). Theorem 14 and Lemma 15 show the following, which may be useful by dramatically reducing the number of constraints required for integrality.[11]

Theorem 16. *If a model has signed ∇G that is odd-K_5-minor-free, then integrality of LP+TRI will be achieved by enforcing only the signed cycle inequalities* (10)*, with the other unsigned cycle inequalities* (9) *being redundant.*[12]

6 DISCUSSION, FUTURE WORK

We have drawn connections to powerful results from graph theory by showing how tightness of LP relaxations may be elegantly characterized by forbidding certain minors: either from the graph topology G, if singleton potentials are not examined; or, with more precision, from the suspension graph ∇G, if the topology of both edge and singleton potentials is considered. We significantly strengthen results by examining also the signs of the potentials and forbidding signed minors. All conditions can be tested efficiently (Theorems 2 and 3). Our strongest result, Theorem 14, shows that LP+TRI is tight for all valid potentials, observing the signed topology of the suspension graph ∇G, iff signed ∇G is odd-K_5-minor-free. Our results go substantially beyond earlier work (Weller et al., 2016) that provided only sufficient conditions for a smaller set of models, without an easy way to test.

Viewing our characterizations together in Table 2, fascinating patterns emerge. We make the following observations. (a) In all known cases, it is exactly just the *odd* versions of the forbidden unsigned minors which can cause lack of tightness of the LP relaxation. In future work, we would like to understand if this pattern extends to other cases. (b) For unsigned graphs G, given the treewidth result of Wainwright and Jordan (2004), we can expect that as cluster size increases, the number of forbidden minors could grow rapidly,[13] see Table 1. (c) For TRI, going from G to the suspension graph ∇G adds one universally connected vertex to the forbidden minor. Why does this not happen for LOC, and will it hold for higher cluster sizes? Recall the observations just before §5.2, where we saw that LOC has a *fixed root* whereas TRI is *universally rooted*. This is why results for TRI examine the suspension graph ∇G with complete symmetry for singleton and edge potentials, whereas for LOC, singleton potentials are different. This prompts further analysis and may lead to new algorithms for TRI. It also suggests viewing the forbidden K_5 minor in ∇G not as a strengthened form of planarity (which forbids K_5 and $K_{3,3}$), but rather as forbidding K_{4+1}, where K_4 is the treewidth constraint of the base graph G. Further, it explains why it is possible for LP+TRI to perform worse as singleton potentials rise within a range, see Appendix.

Theorem 7 shows that for any cluster size r, it is *necessary* to forbid K_{r+2} as a signed minor of ∇G in order to guarantee tightness of LP+$\mathcal{L}_r$. We have placed the appropriate entries in the $\mathcal{L}_4$ row of Table 2. Theorem 5 shows that it is *sufficient* to forbid all the treewidth minors in unsigned G. Must we forbid odd versions of all these in signed G? So far, we have not been able to find an example where LP+$\mathcal{L}_4$ is not tight other than where ∇G contains an odd-K_6.

Acknowledgements

The author thanks Standa Živný, Johannes Carmesin, Maria Chudnovsky, Mark Rowland, Tony Huynh, Tony Jebara, Amir Globerson, David Sontag, Nilesh Tripuraneni, Yarin Gal and Yingzhen Li for helpful comments.

[11] For implementations which successively add violated cutting planes, this result may be less useful, though it still dramatically reduces the search space of possible constraints to add.

[12] Consider that (9) has no access to edge signs, hence tests *all possible* frustrated/odd cycles (10).

[13] Ramachandramurthi (1997) showed that the number of forbidden minors for treewidth t is $\Omega(\exp(\sqrt{t}))$.

References

F. Barahona. The max-cut problem on graphs not contractible to K_5. *Operations Research Letters*, 2(3):107–111, 1983.

F. Barahona. On cuts and matchings in planar graphs. *Math. Program.*, 60:53–68, 1993.

F. Barahona and A. Mahjoub. On the cut polytope. *Mathematical Programming*, 36(2):157–173, 1986.

D. Batra, S. Nowozin, and P. Kohli. Tighter relaxations for MAP-MRF inference: A local primal-dual gap based separation algorithm. In *AISTATS*, 2011.

V. Chandrasekaran, N. Srebro, and P. Harsha. Complexity of inference in graphical models. In *UAI*, pages 70–78, 2008.

M. Cooper and S. Živný. Hybrid tractability of valued constraint problems. *Artificial Intelligence*, 175(9):1555–1569, 2011.

R. Cowell, A. Dawid, S. Lauritzen, and D. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag, 1999.

C. De Simone. The cut polytope and the Boolean quadric polytope. *Discrete Mathematics*, 79(1):71–75, 1990.

E. Demaine, M. Hajiaghayi, and K. Kawarabayashi. Decomposition, approximation, and coloring of odd-minor-free graphs. In *ACM-SIAM symposium on Discrete Algorithms (SODA)*, 2010.

M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer Publishing Company, Incorporated, 1st edition, 1997. ISBN 978-3-642-04294-2.

R. Diestel. *Graph Theory*. Springer, fourth edition, 2010.

J. Domke. Learning graphical model parameters with approximate marginal inference. *TPAMI*, 35(10):2454–2467, 2013.

F. Eaton and Z. Ghahramani. Model reductions for inference: Generality of pairwise, binary, and planar factor graphs. *Neural Computation*, 25(5):1213–1260, 2013.

J. Geelen, B. Gerards, and G. Whittle. Solving Rota's conjecture. *Notices of the AMS*, 61(7):736–743, 2014.

A. Gerards. *Graphs and polyhedra: binary spaces and cutting planes*. 1988.

B. Guenin. A characterization of weakly bipartite graphs. *Journal of Combinatorial Theory, Series B*, 83(1):112–168, 2001.

F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2:143–146, 1953.

T. Huynh. *The linkage problem for group-labelled graphs*. PhD thesis, University of Waterloo, 2009.

J. Kappes, B. Andres, F. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. Kausler, J. Lellmann, N. Komodakis, and C. Rother. A comparative study of modern inference techniques for discrete energy minimization problems. In *CVPR*, 2013.

K. Kawarabayashi, Y. Kobayashi, and B. Reed. The disjoint paths problem in quadratic time. *Journal of Combinatorial Theory, Series B*, 102(2):424–435, 2012.

V. Kolmogorov, J. Thapper, and S. Živný. The power of linear programming for general-valued CSPs. *SIAM Journal on Computing*, 44(1):1–36, 2015. doi: 10.1137/130945648.

J. Kwisthout, H. Bodlaender, and L. van der Gaag. The necessity of bounded treewidth for efficient inference in Bayesian networks. In *ECAI*, volume 215, pages 237–242, 2010.

A. Lehman. On the width-length inequality and degenerate projective planes. In *Polyhedral Combinatorics*, pages 101–105. American Mathematical Society, 1990.

M. Padberg. The Boolean quadric polytope: some characteristics, facets and relatives. *Mathematical Programming*, 45(1-3): 139–172, 1989.

S. Ramachandramurthi. The structure and number of obstructions to treewidth. *SIAM Journal on Discrete Mathematics*, 10(1): 146–157, 1997.

N. Robertson and P. Seymour. Graph minors. XIII. The disjoint paths problem. *Journal of Combinatorial Theory, Series B*, 63 (1):65–110, 1995.

N. Robertson and P. Seymour. Graph minors. XX. Wagner's conjecture. *Journal of Combinatorial Theory, Series B*, 92(2):325–357, 2004.

N. Robertson, P. Seymour, and R. Thomas. Quickly excluding a planar graph. *Journal of Combinatorial Theory, Series B*, 62 (2):323–348, 1994.

T. Schiex, H. Fargier, and G. Verfaillie. Valued constraint satisfaction problems: Hard and easy problems. *IJCAI (1)*, 95: 631–639, 1995.

A. Schrijver. A short proof of Guenin's characterization of weakly bipartite graphs. *Journal of Combinatorial Theory, Series B*, 85 (2):255–260, 2002.

P. Seymour. The matroids with the max-flow min-cut property. *Journal of Combinatorial Theory, Series B*, 23(2):189–222, 1977.

P. Seymour. The origin of the notion of treewidth. `http://cstheory.stackexchange.com/questions/5018/the-origin-of-the-notion-of-treewidth`, 2014.

H. Sherali and W. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

D. Sontag. *Approximate Inference in Graphical Models using LP Relaxations*. PhD thesis, MIT, EECS, 2010.

J. Thapper and S. Živný. The complexity of finite-valued CSPs. Technical report, February 2015. arXiv:1210.2987v3.

K. Truemper. *Matroid decomposition (revised edition)*. 1998.

K. Wagner. Über eine erweiterung eines satzes von Kuratowski. *Deutsche Mathematik*, 2:280–285, 1937.

M. Wainwright and M. Jordan. Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations. *Univ. California, Berkeley, Technical Report*, 671:4, 2004.

M. Wainwright and M. Jordan. Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Y. Watanabe. Uniqueness of belief propagation on signed graphs. In *Neural Information Processing Systems*, 2011.

A. Weller. Revisiting the limits of MAP inference by MWSS on perfect graphs. In *Artificial Intelligence and Statistics (AISTATS)*, 2015.

A. Weller. Uprooting and rerooting graphical models. In *International Conference on Machine Learning (ICML)*, 2016.

A. Weller, M. Rowland, and D. Sontag. Tightness of LP relaxations for almost balanced models. In *Artificial Intelligence and Statistics (AISTATS)*, 2016.

APPENDIX: SUPPLEMENTARY MATERIAL
Characterizing Tightness of LP Relaxations by Forbidding Signed Minors

Here we expand on the observation in §6 of the main paper that LP+TRI can sometimes perform worse as singleton potential strengths increase. We also observe the power of Theorem 14 beyond Theorems 10 or 13.

It is easily seen that if all singleton potentials are sufficiently strong relative to edge potentials, then for each variable, local considerations will be sufficient to determine its MAP configuration, and hence any LP+$\mathcal{L}_r$ relaxation (even on LOC) will be tight. For LP+LOC: if there is a frustrated cycle then stronger edge potentials typically lead to a worse approximation; stronger singleton potentials typically lead to a better approximation.

By examining the semimetric polytope MET, which is equivalent to TRI, we noted just before §5.2 that TRI may be considered *universally rooted*, and hence, in contrast to LOC, TRI treats singleton and edge potentials in an elegantly symmetric way.

In Figure 5, we provide an illustration of this effect. We consider the fully connected model on 4 variables with $W_{ij} = -1 \; \forall (i,j) \in E$ as all θ_i potentials are varied together. The signed graph G is an odd-K_4. From Theorem 14, we know that LP+TRI will be tight if there is no odd-K_5 minor in ∇G. In particular, it must be tight if all singleton potentials are 0 (it must also be tight if any of the singleton potentials were to take opposite signs since this avoids an odd-K_5; this demonstrates the power of Theorem 14 beyond Theorem 10 or Theorem 13).

As θ_i moves away from 0, we obtain an odd-K_5 and the error of LP+TRI increases. Note that the error is symmetric for θ_i on either side of 0, which may be understood by seeing that in the signed suspension graph ∇G, we may resign by flipping the 0 node, which exactly flips the signs of all θ_i potentials; see 3.1.

Developing a better understanding of how approximation error varies with potential strengths is a promising area for future research.

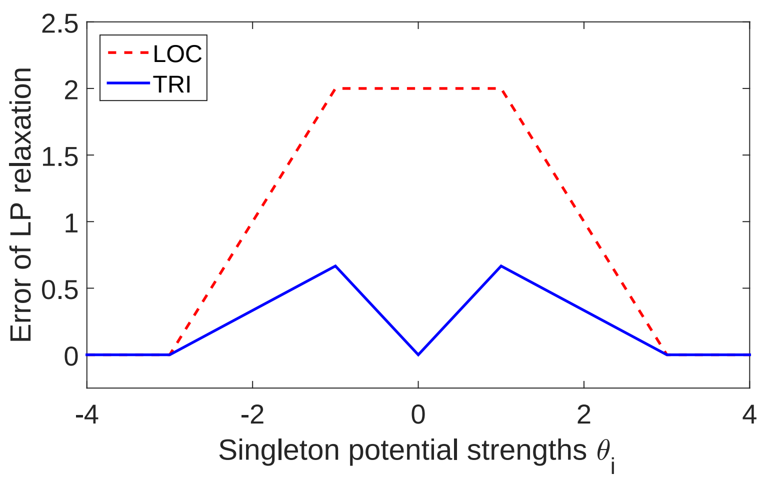


Figure 5: Error of the LP+LOC and LP+TRI relaxations (optimum score minus the optimum integral score) for a fully connected model on 4 variables with $W_{ij} = -1 \; \forall (i,j) \in E$. All θ_i singleton potentials take the same value, which is varied as shown. For sufficiently strong singleton potentials, both LP+LOC and LP+TRI are tight. Starting from 0 singleton potentials, LP+LOC performs monotonically better as singleton potentials strengthen. LP+TRI, in contrast, is tight for 0 singleton potentials, then gets worse and then better again as singleton potentials strengthen.