Conditions Beyond Treewidth for Tightness of Higher-order LP Relaxations

Mark Rowland
University of Cambridge

Aldo Pacchiano
UC Berkeley

Adrian Weller
University of Cambridge

Abstract

Linear programming (LP) relaxations are a popular method to attempt to find a most likely configuration of a discrete graphical model. If a solution to the relaxed problem is obtained at an integral vertex then the solution is guaranteed to be exact and we say that the relaxation is tight. We consider binary pairwise models and introduce new methods which allow us to demonstrate refined conditions for tightness of LP relaxations in the Sherali-Adams hierarchy. Our results include showing that for higher order LP relaxations, treewidth is not precisely the right way to characterize tightness. This work is primarily theoretical, with insights that can improve efficiency in practice.

1 INTRODUCTION

Discrete undirected graphical models are widely used in machine learning, providing a powerful and compact way to model relationships between variables. A key challenge is to identify a most likely configuration of variables, termed *maximum a posteriori* (MAP) or *most probable explanation* (MPE) inference. There is an extensive literature on this problem from various communities, where it may be described as energy minimization (Kappes et al., 2013) or solving a valued constraint satisfaction problem (VCSP, Schiex et al., 1995).

Throughout this paper, we focus on the important class of binary pairwise models (Ising models), allowing arbitrary singleton and edge potentials. For this class, the MAP problem is sometimes described as quadratic pseudo-Boolean optimization (QPBO, e.g. Hammer et al., 1984). In these models, each edge potential may be characterized as either *attractive* (tending to pull its end variables toward the same value; equivalent to a submodular cost function) or *repulsive*. Eaton and Ghahramani (2013) showed that any discrete model may be arbitrarily well approximated by a binary pairwise model, though this may require a large increase in the number of variables.

MAP inference is NP-hard for a general binary pairwise model, hence much work has attempted to identify settings where polynomial-time methods are feasible. We call such settings *tractable* and the methods *efficient*.

In this work, we consider a popular approach which first expresses the MAP problem as an integer linear program (ILP) then relaxes this to a linear program (LP), see §2 for details. An LP attains an optimum at a vertex of the feasible region: if the 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 the LP will always be tight but exponentially many constraints are required, hence the method is not efficient. The marginal polytope $\mathbb{M}$ is typically relaxed to the *local polytope* $\mathbb{L}_2$, which enforces consistency only over each pair of variables (thus yielding *pseudomarginals* which are pairwise consistent but may not correspond to any true global distribution), requiring only a number of constraints which is linear in the number of edges.

LP relaxations are widely used in practice. However, the most common form LP+$\mathbb{L}_2$ often yields a fractional solution, thus motivating more accurate approaches which enforce higher order cluster consistency (Batra et al., 2011). A well-studied example is foreground-background image segmentation. If edge potentials are learned from data, because objects in the real world are contiguous, most edges will be attractive (typically neighboring pixels will be pulled toward the same identification of foreground or background unless there is strong local data from color or intensity). On the horses dataset considered by Domke (2013), LP+$\mathbb{L}_2$ is loose but if triplet constraints are added, then the LP relaxation is often tight. Our work helps to explain and understand such phenomena. This has clear theoretical value and can improve efficiency in practice.

Sherali and Adams (1990) introduced a series of successively tighter relaxations of the marginal polytope: for any integer r, $\mathbb{L}_r$ enforces consistency over all clusters of vari-

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

ables of size $\leq r$. For any fixed r, LP+$\mathbb{L}_r$ is solvable in polynomial time: higher r leads to improved accuracy but higher runtime. Most earlier work considered the case $r = 2$, though recently there has been progress in understanding conditions for tightness for LP+$\mathbb{L}_3$ (Weller et al., 2016; Weller, 2016b,a).

Here we significantly improve on the result for $\mathbb{L}_3$ of Weller et al. (2016), and provide important new results for when LP+$\mathbb{L}_4$ is guaranteed to be tight, employing an interesting geometric perspective. Our main contributions are summarized in §1.3. We first develop the background and context, see (Deza and Laurent, 1997) for a more extensive survey.

Most previous work considers separately two different types of restricted settings that guarantee tightness, either: (i) constraining the potentials to particular families; or (ii) placing structural restrictions on the topology of connections between variables. As an example of the first type of restriction, it is known that if all edge potentials are attractive (equivalently, if all cost functions are submodular), then the basic relaxation LP+$\mathbb{L}_2$ is tight. In fact, Thapper and Živný (2016) showed that for discrete models with variables with any finite number of labels, and potentials of any order: if no restriction is placed on topology, then for a given family of potentials, either the LP relaxation on the natural local polytope is always tight, and hence solves all such problems efficiently, or the problem set is NP-hard.

1.1 Treewidth, Minors and Conditions for Tightness

Exploring the class of structural restrictions, Chandrasekaran et al. (2008) showed that subject to mild assumptions, if no restriction is placed on types of potentials, then the structural constraint of *bounded treewidth* is needed for tractable *marginal* inference.[1] Indeed, Wainwright and Jordan (2004) proved that if a model topology has treewidth $\leq r - 1$ then this is sufficient to guarantee tightness for LP+$\mathbb{L}_r$. As a well-known simple example, if a connected model has treewidth 1 (equivalent to being a tree), then the standard relaxation LP+$\mathbb{L}_2$ is always tight.

The graph property of treewidth $\leq r-1$ is *closed under taking minors* (definitions in §5.1, examples in Figures 1 and 2), hence by the celebrated *graph minor theorem* (Robertson and Seymour, 2004), the property may be characterized by forbidding a unique finite set of minimal *forbidden minors*. Said differently, of all the graphs with treewidth $\geq r$, there is a unique finite set T_r of graphs which are minimal with respect to minor operations. Hence, the sufficient condition of Wainwright and Jordan (2004) for tightness of LP+$\mathbb{L}_r$ may be reframed as: if the graph of a model does not contain any graph in the set T_r as a minor, then LP+$\mathbb{L}_r$ is guaranteed to be tight for any potentials.

The relevant sets of forbidden minors for $\mathbb{L}_2$ and $\mathbb{L}_3$ are particularly simple with just one member each: $T_2 = \{K_3\}$ and $T_3 = \{K_4\}$, where K_n is the complete graph on n vertices. For higher values of r, T_r always contains K_r but there are also other forbidden minors, and their number grows rapidly: T_4 has 4 members (Arnborg et al., 1990) while T_5 has over 70 (Sanders, 1993).

Weller (2016a) showed that, for any r, the graph property of LP+$\mathbb{L}_r$ being *tight for all valid potentials* on the graph is also closed under taking minors. Hence, by Robertson-Seymour, the property for LP+$\mathbb{L}_r$ may be characterized by forbidding a unique set of minimal forbidden minors U_r. It was shown that, in fact, $U_2 = T_2 = \{K_3\}$ and $U_3 = T_3 = \{K_4\}$. However, until this work, all that was known about U_4 is that it contains the complete graph K_5: it has been an open question whether or not $U_4 = T_4$.

One of our main contributions here is to show that $U_4 \neq T_4$. Indeed, in §5 we show that $U_4 \cap T_4 = \{K_5\}$ and that U_4 must contain at least one other forbidden minor, which we cannot yet identify. This progress on understanding U_4 is a significant theoretical development, demonstrating that in general, treewidth is not precisely the right way to characterize tightness of LP relaxations.

Whereas Weller (2016a) made extensive use of powerful earlier results in combinatorics in order to identify U_3, including two results which won the prestigious Fulkerson prize (Lehman, 1990; Guenin, 2001), our analysis takes a different, geometric approach (developed in §4 and §5), which may be of independent interest.

1.2 Stronger Hybrid Conditions

Throughout §1.1, we considered only the graph topology of a model's edge potentials. If we also have access to the *signs* of each edge (attractive or repulsive), then stronger results may be derived. By combining restrictions on both classes of potentials *and* structure, these are termed *hybrid* conditions (Cooper and Živný, 2011).

In this direction, Weller (2016a) showed that for a signed graph, the property that LP+$\mathbb{L}_r$ is tight for all valid potentials (now respecting the graph structure *and* edge signs), is again closed under taking minors, hence again may be characterized by a finite set of minimal forbidden minors U'_r. Further, Weller showed that for $r = 2$ and $r = 3$, the forbidden minors are precisely only the *odd* versions of the forbidden minors for a standard unsigned graph, where an odd version of a graph G means the signed version of G where every edge is repulsive (a repulsive edge is sometimes called odd). That is, $U'_2 = \{\text{odd-}K_3\}$ and $U'_3 = \{\text{odd-}K_4\}$. To see the increased power of these results, observe for example that this means that LP+$\mathbb{L}_3$ is

[1] The treewidth of a graph is one less than the minimum size of a maximum clique in a triangulation of the graph, as used in the junction tree algorithm. Marginal inference seeks the marginal probability distribution for a subset of variables, which is typically harder than MAP inference.

tight for any model of any treewidth, even if it contains K_4 minors, provided only that it does not contain the particular signing of K_4 where all edges are odd (or an equivalent resigning thereof, see §5.1).

In §5, we show somewhat similarly that of all possible signings of K_5, it is only an odd-K_5 which leads to non-tightness of LP+$\mathbb{L}_4$.

1.3 Main Contributions

Given the background in §1.1 and §1.2, here we highlight key contributions.

In §3, we significantly strengthen the result of Weller et al. (2016) for an *almost balanced* model (which contains a distinguished variable s.t. removing it renders the model balanced, see §3.1 for precise definitions). For such a model, it was shown that LP+$\mathbb{L}_3$ is always tight. Here we show that we may relax the polytope from $\mathbb{L}_3$ to a variant we call $\mathbb{L}_3^s$, which enforces triplet constraints only for those triplets including the one distinguished variable s, while still guaranteeing tightness. This has important practical implications since we are guaranteed tightness with dramatically fewer linear constraints, and thus much faster runtime.

We also show in §3 that for any model (no restriction on potentials), enforcing all triplet constraints of $\mathbb{L}_3$ is equivalent to enforcing only those involving the edges of any triangulation (chordal envelope) of its graph, which may significantly improve runtime.

In §4, we introduce a geometric perspective on the tightness of LP relaxations, which may be of independent interest.

In §5, we use these geometric methods to provide powerful new conditions on the tightness of LP+$\mathbb{L}_4$. These show that the relationship which holds between forbidden minors characterizing treewidth and LP+$\mathbb{L}_r$ tightness for $r = 2$ and $r = 3$ breaks down for $r = 4$, hence demonstrating that treewidth is not precisely the right condition for analyzing tightness of higher-order LP relaxations.

1.4 Related Work

We discuss related work throughout the text. To our knowledge, aside from (Weller, 2016a), there is little prior work which considers conditions on signed minors for inference in graphical models. Watanabe (2011) derives a similar characterization to identify when belief propagation has a unique fixed point.

2 BACKGROUND AND PRELIMINARIES

2.1 The MAP Inference Problem

A binary pairwise graphical model is a collection of random variables $(X_i)_{i \in V}$, each taking values in $\{0, 1\}$, such that the joint probability distribution may be written in a minimal representation (Wainwright and Jordan, 2008) as

$$\mathbb{P}(X_i = x_i \; \forall i \in V) \propto \exp\left(\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_i x_j\right), \tag{1}$$

for some potentials $\theta_i \in \mathbb{R}$ for all $i \in V$, and $W_{ij} \in \mathbb{R}$ for all $ij \in E \subseteq V^{(2)}$. We identify the topology of the model as the graph $G = (V, E)$. When $W_{ij} > 0$, there is a preference for X_i and X_j to take the same setting and the edge ij is *attractive*; when $W_{ij} < 0$, the edge is *repulsive* (see Weller, 2014, §2 for details).

A fundamental problem for graphical models is *maximum a posteriori* (MAP) inference, which asks for the identification of a most likely joint state of all the random variables $(X_i)_{i \in V}$ under the probability distribution specified in (1).

The MAP inference problem is clearly equivalent to maximizing the argument of the exponential in (1), yielding the following integer quadratic program:

$$\max_{x \in \{0,1\}^V} \left[\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_i x_j\right]. \tag{2}$$

2.2 LP Relaxations for MAP Inference

A widely used approach to solving Problem (2) is first to replace the integer programming problem with an equivalent linear program (LP), and then to optimize the objective over a relaxed polytope with a polynomial number of linear constraints. This leads to an LP which is efficient to solve but may return a fractional solution (in which case, branching or cutting approaches are often used in practice).

In detail, from Problem 2, an auxiliary variable $x_{ij} = x_i x_j$ is introduced for each edge, as in Problem (3):

$$\max_{\substack{x \in \{0,1\}^{V \cup E} \\ x_{ij} = x_i x_j \forall ij \in E}} \left[\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_{ij}\right]. \tag{3}$$

With the objective now linear, an LP may be formed by optimizing over the convex hull of the optimization domain of Problem (3). This convex hull is called the *marginal polytope* (Wainwright and Jordan, 2008) and denoted $\mathbb{M}(G)$. Thus we obtain the equivalent problem:

$$\max_{q \in \mathbb{M}(G)} \left[\sum_{i \in V} \theta_i q_i + \sum_{ij \in E} W_{ij} q_{ij}\right]. \tag{4}$$

This LP (4) is in general no more tractable than Problem (2), due to the number of constraints needed to describe $\mathbb{M}(G)$. It is standard to obtain a tractable problem by *relaxing* the domain of optimization $\mathbb{M}(G)$ to some larger polytope $\mathbb{L}$ which is easier to describe. The resulting relaxed

optimization yields an upper bound on the optimal value of Problem (2), though if the $\arg\max$ over $\mathbb{L}$ is an extremal point of $\mathbb{M}(G)$, then the approximation must be exact, and we say that the relaxation is *tight* for this problem instance.

We focus on the family of relaxations introduced by Sherali and Adams (1990), defined below. We first consider a probabilistic interpretation of the marginal polytope.

Notation. For any finite set Z, let $\mathscr{P}(Z)$ be the set of probability measures on Z.

The marginal polytope can then be written

$$\begin{aligned}\mathbb{M}(G) = \{((q_i)_{i\in V}, (q_{ij})_{ij\in E}) \mid \exists \mu \in \mathscr{P}(\{0,1\}^V) \\ \text{s.t. } q_i = \mathbb{P}_\mu(X_i = 1)\ \forall i \in V\,, \\ q_{ij} = \mathbb{P}_\mu(X_i = 1, X_j = 1)\ \forall ij \in E\}.\end{aligned} \tag{5}$$

Intuitively the constraints of the marginal polytope enforce a global consistency condition for the set of parameters $((q_i)_{i\in V}, (q_{ij})_{ij\in E})$, so that together they describe marginal distributions of some global distribution over the entire set of variables. A natural approach is to relax this condition of global consistency to a less stringent notion of consistency only for smaller clusters of variables.

Definition 1. The Sherali-Adams polytope $\mathbb{L}_r(G)$ of order r for a binary pairwise graphical model on $G = (V, E)$ is

$$\begin{aligned}\mathbb{L}_r(G) = \{((q_i)_{i\in V}, (q_{ij})_{ij\in E}) \mid \forall \alpha \subseteq V \text{ with } |\alpha| \le r, \\ \exists \tau_\alpha \in \mathscr{P}(\{0,1\}^\alpha) \text{ s.t. } \tau_\alpha\big|_\beta = \tau_\beta\ \forall \beta \subset \alpha, \text{ and} \\ \tau_{\{i,j\}}(1,1) = q_{ij}\ \forall ij \in E,\ \tau_{\{i\}}(1) = q_i\ \forall i \in V\},\end{aligned} \tag{6}$$

where for $\beta \subset \alpha$, $\tau_\alpha\big|_\beta$ denotes the marginal distribution of τ_α on $\{0,1\}^\beta$.

Considering successive fixed values of $r \in \mathbb{N}$, the Sherali-Adams polytopes therefore yield the following sequence of tractable approximations to Problem (2):

$$\max_{q\in\mathbb{L}_r(G)} \left[\sum_{i\in V} \theta_i q_i + \sum_{ij\in E} W_{ij} q_{ij}\right]. \tag{7}$$

As r is increased, so too does the number of linear constraints required to define $\mathbb{L}_r$, leading to a tighter polytope and a more accurate solution, but at the cost of greater computationally complexity. For $r = |V|$, $\mathbb{L}_r(G)$ is exactly equal to the marginal polytope $\mathbb{M}(G)$.

A fundamental question concerning LP relaxations for MAP inference is when it is possible to use a computationally cheap relaxation $\mathbb{L}_r(G)$ and still obtain an exact answer to the original inference problem. This is of great practical importance, as it leads to tractable algorithms for particular classes of problems that in full generality are NP-hard. In this paper we investigate this question for a variety of problem classes for the Sherali-Adams relaxations: $\mathbb{L}_2(G)$, $\mathbb{L}_3(G)$, and $\mathbb{L}_4(G)$; also referred to as the local, triplet, and quad polytopes respectively.

3 REFINING TIGHTNESS RESULTS FOR $\mathbb{L}_3(G)$

Here we first derive new results for the triplet polytope $\mathbb{L}_3$ that significantly strengthen earlier work (Weller et al., 2016) for *almost balanced* models. Then in §3.3, we demonstrate that for any model, triplet constraints need be applied only over edges in some triangulation of its graph (that is, over any chordal envelope).

3.1 Graph-theoretic Preliminaries

A *signed graph* is a graph $G = (V, E)$ together with a function $\Sigma : E \to \{\text{even, odd}\}$, where attractive edges are even and repulsive edges are odd. A signed graph is *balanced* (Harary, 1953) if its vertices V may be partitioned into two exhaustive sets V_1, V_2 such that all edges with both endpoints in V_1 or both endpoints in V_2 are even, whilst all edges with one endpoint in each of V_1 and V_2 is odd.

A signed graph $G = (V, E, \Sigma)$ is *almost balanced* (Weller, 2015b) if there exists some distinguished vertex $s \in V$ such that removing s leaves the remainder balanced (thus any balanced graph is almost balanced). To detect if a graph is almost balanced, and if so then to find a distinguished vertex, may be performed efficiently (simply hold out one variable at a time and test the remainder to see if it is balanced, Harary and Kabell, 1980). A graphical model is almost balanced if the signed graph corresponding to its edge potentials is almost balanced.

3.2 Tightness of LP Relaxations for Almost Balanced Models

The following result was shown by Weller et al. (2016).

Theorem 2. *Given a graph G, the triplet polytope $\mathbb{L}_3(G)$ is tight on the class of almost balanced models on G.*

We present a significant strengthening of Theorem 2 in a new direction, which identifies exactly which constraints of $\mathbb{L}_3(G)$ are required to ensure tightness for the class of almost balanced models.

Given an almost balanced model with distinguished variable s (that is, deleting s renders the model balanced), define the polytope $\mathbb{L}_3^s(G)$ by taking all the pairwise constraints of $\mathbb{L}_2(G)$, and adding triplet constraints *only* for triplets of variables that include s (see §10.4.1 in the Supplement for more details).

$\mathbb{L}_3^s(G)$ is a significant relaxation of $\mathbb{L}_3(G)$, requiring only $\mathcal{O}(|V|^2)$ linear constraints, rather than the $\mathcal{O}(|V|^3)$ constraints needed for $\mathbb{L}_3(G)$. Thus it is substantially faster to optimize over $\mathbb{L}_3^s(G)$. Nevertheless, our next result shows

that $\mathbb{L}_3^s(G)$ is still tight for an almost balanced model, and indeed is the 'loosest' possible polytope with this property.

Theorem 3. *The polytope $\mathbb{L}_3^s(G)$ is tight on the class of almost balanced models on graph G with distinguished variable s. Further, no linear constraint of $\mathbb{L}_3^s(G)$ may be removed to yield a polytope which is still tight on all models in this class of potentials. Proof in the Supplement §10.*

Considering cutting plane methods, Theorem 3 demonstrates exactly which constraints from $\mathbb{L}_3(G)$ may be necessary to add to the polytope $\mathbb{L}_2(G)$ in order to achieve tightness for an almost balanced model.

3.3 Chordality and Extending Partial Marginals

By its definition, the polytope $\mathbb{L}_2$ enforces pairwise constraints on every pair of variables in a model, whether or not they are connected by an edge, yet typically one enforces constraints only for edges E in the model. This is sufficient because it is not hard to see that if one has edge marginals for the graph $G(V, E)$, it is always possible to extend these to edge marginals for the complete graph on V while remaining within $\mathbb{L}_2$ (and the values of these additional marginals are irrelevant to the score by assumption).

Here we provide an analogous result for $\mathbb{L}_3$, which shows that for any model (no restriction on potentials), one need only enforce triplet constraints over any triangulation (chordal envelope) of its graph.

Theorem 4. *For a chordal graph G, the polytope $\mathbb{L}_2(G)$ together with the triplet constraints only for those triplets of variables that form 3-cliques in G, is equal to $\mathbb{L}_3(G)$, i.e. the polytope given by enforcing constraints on all triplets of G. Proof in the Supplement §9.*

4 THE GEOMETRY OF SHERALI-ADAMS RELAXATIONS

Here we introduce several geometric notions for the Sherali-Adams polytopes which we shall apply in §5.

4.1 The Geometry of the Sherali-Adams Polytopes

The study of tightness of LP relaxations is naturally expressed in the language of polyhedral combinatorics. We introduce key notions from the literature, then provide new proofs of characterizations of tightness for the local polytope $\mathbb{L}_2(G)$ with these geometric ideas.

Given a polytope $P \subset \mathbb{R}^m$, and an extremal point (vertex) $v \in \mathrm{Ext}(P)$, the *normal cone* to P at v, denoted $N_P(v)$, is the polyhedral cone defined by

$$N_P(v) = \left\{ c \in \mathbb{R}^m | v \in \arg\max_{x \in P} \langle c, x \rangle \right\}. \tag{8}$$

We define the *conical hull* of a finite set X as: $\mathrm{Cone}(X) = \{\sum_{x \in X} \lambda_x x \mid \lambda_x \geq 0 \ \forall x \in X\}$. The following characterization of normal cones will be particularly useful.

Lemma 5 (Theorem 2.4.9, Schneider, 1993)**.** *Let $P = \{x \in \mathbb{R}^m | Ax \leq b\}$ be a polytope for some $A = [a_1, \ldots, a_k]^\top \in \mathbb{R}^{k \times m}$, $b \in \mathbb{R}^k$ (for some $k \in \mathbb{N}$). Then for $v \in \mathrm{Ext}(P)$, we have*

$$N_P(v) = \mathrm{Cone}(\{a_i | \langle a_i, v \rangle = b\}). \tag{9}$$

Further, if the representation $\{x \in \mathbb{R}^m | Ax \leq b\}$ has no redundant constraints, then $\{a_i | \langle a_i, v \rangle = b\}$ is a complete set of extremal rays of $N_P(v)$ (up to scalar multiplication).

With these geometric notions in hand, we have a succinct characterization of the set of potentials for which a given Sherali-Adams relaxation is tight.

Lemma 6. *The set of potentials which are tight with respect to $\mathbb{L}_r$ is exactly given by the following union of cones*

$$\bigcup_{v \in \mathrm{Ext}(\mathbb{M}(G))} N_{\mathbb{L}_r(G)}(v). \tag{10}$$

This concise characterization, together with the explicit parametrization of normal cones given by Lemma 5 and the form of the linear constraints defining the local polytope $\mathbb{L}_2(G)$, yields an efficient algorithm for generating arbitrary potentials which are tight with respect to $\mathbb{L}_2(G)$.

We would like also to identify classes of potentials for which $\mathbb{L}_2(G)$ is guaranteed to be tight. We demonstrate the power of our geometric approach by providing new proofs in §7 of the Supplement of the following earlier results.

Lemma 7. *If G is a tree, then $\mathbb{L}_2(G)$ is tight for all potentials, that is $\mathbb{L}_2(G) = \mathbb{M}(G)$.*

Lemma 8. *For an arbitrary graph G, $\mathbb{L}_2(G)$ is tight for the set of balanced models.*

The proofs proceed by explicitly demonstrating that a given potential lies in a cone $N_{\mathbb{L}_2(G)}(v)$ for some vertex v of the marginal polytope, by expressing the potential as a conical combination of the extremal rays of the cone.

4.2 The Symmetry of the Sherali-Adams Polytopes

The Sherali-Adams polytopes have rich symmetries which can be exploited when classifying tightness of LP relaxations using the tools discussed in §4.1. Intuitively, these symmetries arise either by considering relabellings of the vertices of the graph G (permutations), or relabellings of the state space of individual variables (flippings). The key result is that a Sherali-Adams polytope is tight a for potential c iff it is tight for any permutation or flipping of c.

4.2.1 The Permutation Group

Let $\sigma \in S_V$ be a relabeling of the vertices of the graph G. This permutation then induces a bijective map Y_σ :

$\mathbb{L}_r(G) \to \mathbb{L}_r(G)$ (which naturally lifts to a linear map on $\mathbb{R}^{V \cup E}$), given by applying the corresponding relabeling to the components of the pseudomarginal vectors:

$$Y_\sigma((q_i)_{i\in V}, (q_{ij})_{ij\in E}) = ((q_{\sigma(i)})_{i\in V}, (q_{\sigma(i)\sigma(j)})_{ij\in E}) \quad \forall((q_i)_{i\in V}, (q_{ij})_{ij\in E}) \in \mathbb{L}_r(G)\,.$$

The element $\sigma \in S_V$ also naturally induces a linear map on the space of potentials, which is formally the dual space $(\mathbb{R}^{V\cup E})^\star$, although we will frequently identify it with $\mathbb{R}^{V\cup E}$. We denote the map on the space of potentials by $Y_\sigma^\dagger : (\mathbb{R}^{V\cup E})^\star \to (\mathbb{R}^{V\cup E})^\star$, given by

$$Y_\sigma^\dagger((c_i)_{i\in V}, (c_{ij})_{ij\in E}) = ((c_{\sigma(i)})_{i\in V}, (c_{\sigma(i)\sigma(j)})_{ij\in E}) \quad \forall((c_i)_{i\in V}, (c_{ij})_{ij\in E}) \in \mathbb{R}^{V\cup E}\,.$$

The sets $\{Y_\sigma | \sigma \in S_V\}$ and $\{Y_\sigma^\dagger | \sigma \in S_V\}$ obey the group axioms (under the operation of composition), and hence form groups of symmetries on $\mathbb{L}_r(G)$ and $(\mathbb{R}^{V\cup E})^\star$ respectively; they are both naturally isomorphic to S_V.

These symmetry groups form a useful formalism for thinking about tightness of Sherali-Adams relaxations. We provide one such result in this language, proof in §8 of the Supplement.

Lemma 9. *$\mathbb{L}_r(G)$ is tight for a given potential $c \in \mathbb{R}^{V\cup E}$ iff it is tight for all potentials $Y_\sigma^\dagger(c)$, $\sigma \in S_V$.*

4.2.2 The Flipping Group

Whilst the permutation group described above corresponds to permuting the labels of vertices in the graph, it is also useful to consider the effect of permuting the labels of the states of individual variables. In the case of binary models, the label set is $\{0, 1\}$, so permuting labels corresponds to switching $0 \leftrightarrow 1$, which we refer to as 'flipping'. Given a variable $v \in V$, define the affine map $F_{(v)} : \mathbb{R}^{V\cup E} \to \mathbb{R}^{V\cup E}$ which acts on any pseudomarginal $q \in \mathbb{R}^{V\cup E}$ to flip v as follows (see Weller et al., 2016 for details):

$$\big[F_{(v)}(q)\big]_v = 1 - q_v,$$
$$\big[F_{(v)}(q)\big]_{vw} = q_w - q_{vw},\ \forall vw \in E,$$

while $F_{(v)}$ leaves unchanged all other coordinates of q. Note that $F_{(v)}$ restricts to a bijection on $\mathbb{L}_r(G)$. The flipping maps commute and have order 2, hence the group generated by them, $\langle F_{(v)} | v \in V\rangle$, is isomorphic to $\mathbb{Z}_2^{|V|}$. A general element of this group can be thought of as simultaneously flipping a subset $I \subseteq V$ of variables, written as $F_{(I)} : \mathbb{L}_r(G) \to \mathbb{L}_r(G)$.

Flipping a subset of variables $I \subseteq V$ also naturally induces a map $F_{(I)}^\dagger : (\mathbb{R}^{V\cup E})^\star \to (\mathbb{R}^{V\cup E})^\star$ on the space of potentials. We give a full description of this map in §8 of the Supplement. An analogous result to Lemma 9 also holds for the group of flipping symmetries.

Lemma 10. *$\mathbb{L}_r(G)$ is tight for a given potential $c \in \mathbb{R}^{V\cup E}$ iff it is tight for all potentials $F_{(I)}^\dagger(c)$, $I \subseteq V$.*

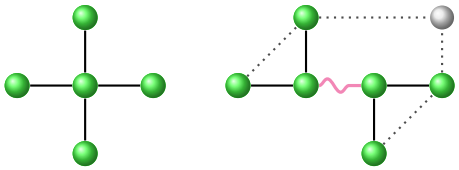


Figure 1: The left graph is a *minor* (unsigned) of the right graph, obtained by deleting the grey dotted edges and resulting isolated grey vertex, and contracting the purple wavy edge. See §5.1.

4.2.3 The Joint Symmetry Group of the Sherali-Adams Polytopes

Tying together the remarks of §4.2.1 and §4.2.2, note that in general the symmetries of the flipping and permutation groups on $\mathbb{L}_r(G)$ do not commute. In fact, observe that

$$Y_\sigma^{-1} \circ F_{(I)} \circ Y_\sigma = F_{(\sigma^{-1}(I))}\,. \tag{11}$$

Thus the group of symmetries of $\mathbb{L}_r(G)$ generated by permutations and flippings is isomorphic to the semidirect product $S_V \rtimes \mathbb{Z}_2^{|V|}$.

5 FORBIDDEN MINOR CONDITIONS FOR TIGHTNESS OF $\mathbb{L}_4(G)$

We first introduce graph minors and their application to the characterizations of both treewidth and tightness of LP relaxations over $\mathbb{L}_r$. While these characterizations are the same for $r \leq 3$, a key contribution in this section is to use the geometric perspective of §4 to show that the characterizations are not the same for $r = 4$.

5.1 Graph Minor Theory

For further background, see (Diestel, 2010, Chapter 12). Given a graph $G = (V, E)$, a graph H is a *minor* of G, written $H \leq G$, if it can be obtained from G via a series of edge deletions, vertex deletions, and edge *contractions*. The result of contracting an edge $uv \in E$ is the graph $G'(V', E')$ where u and v are 'merged' to form a new vertex w which is adjacent to any vertex that was previously adjacent to either u or v. That is $V' = V \setminus \{u, v\} \cup \{w\}$, and $E' = \{e \in E \mid u, v \notin e\} \cup \{wx \mid ux \in E \text{ or } vx \in E\}$. This is illustrated in Figure 1.

A property of a graph is *closed under taking minors* or *minor-closed* if whenever G has the property and $H \leq G$, then H also has the property.

The celebrated *graph minor theorem* of Robertson and Seymour (2004) proves that any minor-closed graph property may be characterized by a unique finite set $\{H_1, \ldots, H_m\}$ of minimal (wrt minor operations) forbidden minors; that is, a graph G has the property iff it does not contain any H_i as a minor. Checking to see if a graph contains some

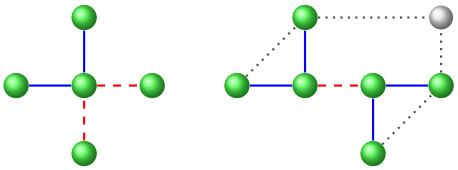


Figure 2: The left graph is a *signed minor* of the right signed graph, obtained similarly to Figure 1 except that before contracting the repulsive edge, first flip the vertex at its right end. Solid blue (dashed red) edges are attractive (repulsive). Grey dotted edges on the right are deleted and may be of any sign. See §5.1.

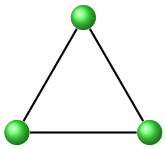


Figure 3: K_3 (unsigned), the only element of T_2, the set of minimal forbidden minors for treewidth ≤ 1. $T_2 = U_2$, the set of forbidden minors for LP+$\mathbb{L}_2$ to be tight. See §5.2.

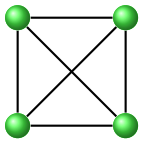
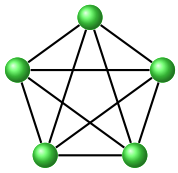
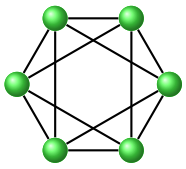


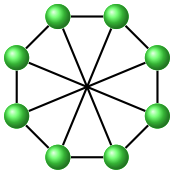
Figure 4: K_4 (unsigned), the only element of T_3, the set of minimal forbidden minors for treewidth ≤ 2. $T_3 = U_3$, the set of forbidden minors for LP+$\mathbb{L}_3$ to be tight. See §5.2.



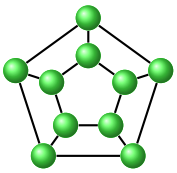
(a) K_5



(b) Octahedral graph



(c) Wagner graph



(d) Pentagonal prism graph

Figure 5: The four members of T_4, the set of minimal forbidden minors (unsigned) for treewidth ≤ 3. We show the new result that $T_4 \neq U_4$, in fact $T_4 \cap U_4 = \{K_5\}$, where U_4 is the set of forbidden minors for LP+$\mathbb{L}_4$ to be tight. See §5.2.

H as a minor may be performed efficiently (Robertson and Seymour, 1995).

We shall also consider signed graphs (see §3.1) and their respective *signed minors*. A signed minor of a signed graph is obtained as before by edge deletion, vertex deletion, and edge contraction but now: contraction may be performed only on even (attractive) edges; and any *resigning* operation is also allowed, in which a subset of vertices $S \subseteq V$ is selected then all edges with exactly one end in S are flipped even $\leftrightarrow$ odd.[2] An example is shown in Figure 2. This notion of flipping is closely related to the notion of flipping of potentials, introduced in §4.2.2.

The graph minor theorem of Robertson and Seymour generalizes to signed graphs (Huynh, 2009; Geelen et al., 2014): any property of a signed graph which is closed under taking signed minors, may be characterized by a unique finite set of minimal forbidden signed minors.

5.2 Treewidth Characterizations of Tightness

A fundamental result in the study of LPs over Sherali-Adams relaxations is the following sufficient condition.

Theorem 11 (Wainwright and Jordan, 2004)**.** *If G has treewidth $\leq r-1$, then $\mathbb{L}_r(G) = \mathbb{M}(G)$; equivalently, LP+$\mathbb{L}_r$ is tight for all valid potentials on G.*

The goal of this section is to study to what extent a partial converse to this result holds; that is, to what extent tightness of a Sherali-Adams polytope can hold for graphs of high treewidth. We focus on the case of $\mathbb{L}_4(G)$, and proceed based on the graph minor properties of G.

See §1.1 for a quick review of results that the properties, for any r, of treewidth$\leq r-1$, and of LP+$\mathbb{L}_r$ being *tight for any valid potentials*, are both minor-closed (Weller, 2016a). Thus, by the graph minor theorem, both are characterized by a finite set of minimal forbidden minors.

We shall often be interested in the set of potentials $\widetilde{\Sigma} := \{c \in \mathbb{R}^{V \cup E} \,|\, \text{sign}(c_e) = \Sigma(e)\ \forall e \in E\}$ that are consistent with a given signing Σ of the graph G. We say that the polytope $\mathbb{L}_r(G)$ is *tight* for a signing Σ if LP+$\mathbb{L}_r$ is tight for all potentials in the set $\widetilde{\Sigma} \subset \mathbb{R}^{V \cup E}$.

The property of a signed graph (G, Σ) that $\mathbb{L}_r(G)$ is tight for the signing Σ is also minor-closed (Weller, 2016a), hence can also be characterized by forbidden minimal signed minors.

Recalling the notation introduced in §1.1: for unsigned graphs, we call the respective sets of minimal forbidden minors T_r (for treewidth $\leq r-1$) and U_r (for LP tightness over $\mathbb{L}_r$); for signed graphs, U'_r is the set of minimal forbidden signed minors for tightness of LP+$\mathbb{L}_r$. It is known (Weller, 2016a) that in fact, $T_r = U_r$ for $r = 2, 3$, and that U'_r is exactly the set of odd signings (i.e. signings where all edges of the graph are odd/repulsive) of the graphs of T_r for $r = 2, 3$. We shall show in §5.3 that both of these relationships break down for $r = 4$.

The sets T_2, T_3 and T_4 are shown in Figures 3 to 5, whilst the sets U'_2 and U'_3 are shown in Figures 6 and 7.

[2] Hence, to contract an odd edge, one may first flip either end of the edge to make it even, then contract. In our context of binary pairwise models, flipping a subset S is equivalent to switching from a model with variables $\{X_i\}$ to a new model with variables $\{Y_i : Y_i = 1 - X_i\ \forall i \in S, Y_i = X_i\ \forall i \in V \setminus S\}$ and setting potentials to preserve the distribution, which flips the sign of W_{ij} for any edge ij with one end in S; details in (Weller, 2015a, §2.4).

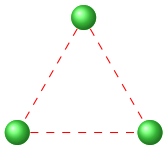


Figure 6: Odd-K_3, the unique element of U_2', the set of minimal forbidden signed minors for tightness of $\mathbb{L}_2(G)$. Red dashed edges represent repulsive edges. See §5.2.

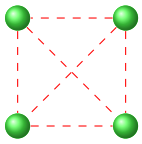


Figure 7: Odd-K_4, the unique element of U_3', the set of minimal forbidden signed minors for tightness of $\mathbb{L}_3(G)$. Red dashed edges represent repulsive edges. See §5.2.

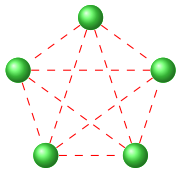


Figure 8: Odd-K_5, the unique signing of an element of T_4, the set of minimal forbidden unsigned minors for treewidth ≤ 3, that appears in U_4', the set of minimal forbidden signed minors for $\mathbb{L}_4(G)$, as shown by our new Theorem 12. It was previously known that for $\mathbb{L}_r, r \leq 3$, the minimal forbidden signed minors for LP-tightness are exactly the odd versions of the minimal forbidden unsigned minors for treewidth $\leq r-1$. We believe there must be at least one other forbidden minor for tightness of $\mathbb{L}_4$, see §5.4. Red dashed edges represent repulsive edges.

5.3 Identifying Forbidden Signed Minors

$\mathbb{L}_4(G)$ is tight for a potential $c \in \mathbb{R}^{V \cup E}$ if and only if

$$\max_{q \in \mathbb{L}_4(G)} \langle c, q \rangle = \max_{x \in \mathbb{M}(G)} \langle c, x \rangle \ , \tag{12}$$

or equivalently if

$$\max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] = 0 \, . \tag{13}$$

Since $\mathbb{M}(G) \subseteq \mathbb{L}_4(G)$, we have $\max_{q \in \mathbb{L}_4(G)} \langle c, q \rangle \geq \max_{x \in \mathbb{M}(G)} \langle c, x \rangle \ \ \forall c \in \mathbb{R}^{V \cup E}$. Hence, it follows that $\mathbb{L}_4(G)$ is not tight for some potential $c \in \widetilde{\Sigma}$ (the set of potentials respecting a signing Σ of G, see §5.2) iff the following optimization problem has a non-zero optimum:

$$\max_{c \in \widetilde{\Sigma}} \max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] \, . \tag{14}$$

For the graphs in T_4, this is a high-dimensional indefinite quadratic program which is intractable to solve. However, using the geometric ideas of §4, we may decompose this problem into a sequence of tractable linear programs. This process involves computing vertex representations (V-representations) for a variety of polytopes using the ideas of §4, and computing the orbits of the set of signings of G under the natural action of the group described in §4.2.3, see the Supplement §11 for full details. Solving these linear programs then allows the exact set of non-tight signings of the graphs in Figure 5 to be identified; see Figure 8.

Theorem 12. *The only non-tight signing for $\mathbb{L}_4$ of any minimal forbidden minor for treewidth ≤ 3 is the odd-K_5.*

5.4 Discussion: Other Forbidden Minors

Previous work showed that tightness for all valid potentials of LP+$\mathbb{L}_2$ may be characterized exactly by forbidding just an odd-K_3 as signed minor, and that a similar result for LP+$\mathbb{L}_3$ holds by forbidding just an odd-K_4 (Weller, 2016a). These are precisely the odd versions of the forbidden minors for the respective treewidth conditions.

A natural conjecture for $\mathbb{L}_4$ was that one must forbid just some signings of the four graphs in T_4, see Figure 5. Now given Theorem 12, it would seem sensible to wonder if LP+$\mathbb{L}_4$ is tight for all valid potentials iff a model's graph does not contain an odd-K_5?

However, this must be false (unless P=NP), since if it were true: We would have LP+$\mathbb{L}_4$ is tight for any model not containing K_5 (as an unsigned minor). It is well-known that planar graphs are those without K_5 or $K_{3,3}$ as a minor ($K_{3,3}$ is the complete bipartite graph where each partition has 3 vertices), i.e. a subclass of graphs which are K_5-free. Hence, we would have a polytime method to solve MAP inference for any planar binary pairwise model. Yet it is not hard to see that we may encode minimal vertex cover in such a model, and it is known that planar minimum vertex cover is NP-hard (Lichtenstein, 1982).

We have not yet been able to identify any other minimal forbidden minor for tightness of LP+$\mathbb{L}_4$, but note that one natural candidate is some signing of a $k \times k$ grid of sufficient size, since this is planar with treewidth k.

6 CONCLUSION

LP relaxations are widely used for the fundamental task of MAP inference for graphical models. Considering binary pairwise models, we have provided important theoretical results on when various relaxations are guaranteed to be tight, which guarantees that in practice, an exact solution can be found efficiently.

A key result focuses on the connection between tightness of LP relaxations of a model and the treewidth of its graph. For the first two levels of the Sherali-Adams hierarchy, that is for the pairwise and triplet relaxations, it was known that the characterizations are essentially identical. However, we have shown that this pattern does not hold for the next level in the hierarchy, that is for the quadruplet polytope $\mathbb{L}_4$.

We refined this result by considering the signed graph of a model and its signed minors. To derive these results we introduced geometric methods which may be of independent interest.

Acknowledgements

We thank Francis Bach, Maria Lomeli and the anonymous reviewers for helpful comments. MR acknowledges support by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L016516/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. AW acknowledges support by the Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

S. Arnborg, A. Proskurowski, and D. Corneil. Forbidden minors characterization of partial 3-trees. *Discrete Mathematics*, 80 (1), February 1990.

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.

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.

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

P. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0–1 optimization. *Mathematical programming*, 28(2):121–155, 1984.

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

F. Harary and J. Kabell. A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences*, 1(1):131–136, 1980.

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.

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

D. Lichtenstein. Planar formulae and their uses. *SIAM Journal on Computing*, 11(2):329–343, 1982.

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.

D. Sanders. *Linear algorithms for graphs of tree-width at most four*. PhD thesis, Georgia Tech, 1993.

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

R. Schneider. *Convex bodies : the Brunn-Minkowski theory*. Cambridge University Press, 1993.

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.

J. Thapper and S. Živný. The complexity of finite-valued CSPs. *Journal of the ACM*, 63(4), 2016.

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. *Methods for Inference in Graphical Models*. PhD thesis, Columbia University, 2014.

A. Weller. Bethe and related pairwise entropy approximations. In *Uncertainty in Artificial Intelligence (UAI)*, 2015a.

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

A. Weller. Characterizing tightness of LP relaxations by forbidding signed minors. In *Uncertainty in Artifical Intelligence (UAI)*, 2016a.

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

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
Conditions Beyond Treewidth for Tightness of Higher-order LP Relaxations

In this Appendix, we provide the following.

- §7 Geometric Proofs of Results for $\mathbb{L}_2$
- §8 Group-theoretic Results from §4.2
- §9 Proof of the Marginals Extension Result for $\mathbb{L}_3$
- §10 Proof of Theorem 3 for $\mathbb{L}_3^s$
- §11 Identifying Non-tight Signings of T_4, Treewidth 3 Forbidden Minors

7 Geometric Proofs of Results for $\mathbb{L}_2$

In this section, we give proofs of Lemmas 7 and 8 using the geometric insights of §4.

Lemma 7. If $G = (V, E)$ is a tree, then $\mathbb{L}_2(G)$ is tight for any binary pairwise model on G.

Proof. We will show that given arbitrary potentials $c \in \mathbb{R}^{V \cup E}$ with MAP optimum configuration $v \in \text{Ext}(\mathbb{M}(G))$, we have $c \in N_{\mathbb{L}_2(G)}(v)$, from which it immediately follows from the definition of normal cones that v is optimal for c in $\mathbb{L}_2(G)$, and hence the result follows.

We first argue that it is sufficient to consider potentials $c \in \mathbb{R}^{V \cup E}$ such that $c \in N_{\mathbb{M}(G)}(v_0)$, where $v_0 \in \{0, 1\}^{V \cup E}$ is the configuration where every random variable X_v in the graphical model is set to 0. To see this, suppose that $c \in \mathbb{R}^{V \cup E} \setminus N_{\mathbb{M}(G)}(v_0)$. Since $\mathbb{R}^{V \cup E} = \cup_{v \in \text{Ext}(\mathbb{M}(G))} N_{\mathbb{M}(G)}(v)$, there exists $v' \in \text{Ext}(\mathbb{M}(G))$ such that $c \in N_{\mathbb{M}(G)}(v')$. Now consider a flipping map F (see §4.1) that maps $v \in \mathbb{M}(G)$ to $v_0 \in \mathbb{M}(G)$, and recall by Lemma 10 that $\mathbb{L}_2(G)$ is tight for c iff it is tight for $F^{\dagger}_{v' \to v_0}(c)$.

We now proceed by induction on the number of vertices n of G. Our method will be to show that $c \in N_{\mathbb{L}_2(G)}(v_0)$ by directly showing that c can be written as a positive linear combination of extremal vectors of $N_{\mathbb{L}_2(G)}(v_0)$. The cases $n = 1, 2$ are trivial, as $\mathbb{L}_2(G) = \mathbb{M}(G)$ for such models immediately from the definition of Sherali-Adams polytopes. For the inductive step, let $x \in G$ be a leaf, with neighbor $y \in G$. As x, y are both equal to 0 at the marginal optimum, there are three tight local constraints (see §10.4.1) corresponding to the pair x, y which are tight at the marginal optimum v_0, namely $q_{xy} \leq q_x$, $q_{xy} \leq q_y$, and $q_{xy} \geq 0$. By Lemma 5, the outward pointing normal to each hyperplane defined by these constraints is an extremal vector of the cone $N_{\mathbb{M}(G)}(v_0)$; using the notation $\mathbf{e}_i$ to denote a unit vector in the direction corresponding to coordinate i, we can write the outward pointing normals as $\mathbf{e}_{xy} - \mathbf{e}_x$, $\mathbf{e}_{xy} - \mathbf{e}_y$, and $-\mathbf{e}_{xy}$. Note that as v_0 is the optimal configuration, setting either of x and y cannot result in a higher scoring configuration, from which we obtain $\theta_x, \theta_y \leq 0$, and $\theta_x + \theta_y + W_{xy} \leq 0$.

We now consider two cases: i) $W_{xy} \geq -\theta_x$; and ii) $W_{xy} < -\theta_x$.

i) If $W_{xy} \geq -\theta_x$, we consider the following conical combination:

$$-\theta_x(\mathbf{e}_{xy} - \mathbf{e}_x) + (W_{xy} + \theta_x)(\mathbf{e}_{xy} - \mathbf{e}_y). \tag{15}$$

Note that the coefficients of $\mathbf{e}_x$ and $\mathbf{e}_{xy}$ exactly match the singleton potential on x and the edge potential y for the model. We now consider the residual potentials (that is, the model whose potentials are the difference between the potentials of the original model, and the potentials given by the conical combination of extremal vectors in (15)); it is sufficient for the residual potentials to have the same optimal marginal vertex for the inductive step to be complete. Note that the residual potentials are the same for all singletons except x and y; the new potential on x is 0 and the new potential on y is $\theta_x + \theta_y + W_{xy}$. If some other vertex yields a higher score than v_0 on this model, it must be the case that y is set to 1 in this model - else the same vertex would yield a greater score than that of v_0 in the original model. But if y is set to 1, the residual model yields the same score as setting both x and y equal to 1 in the original model. Therefore we deduce that v_0 is optimal for the original model, and the inductive step is complete.

ii) If $W_{xy} < -\theta_x$, we consider the following conical combination

$$-\theta_x(\mathbf{e}_{xy} - \mathbf{e}_x) + (-W_{xy} - \theta_x)(-\mathbf{e}_{xy}). \tag{16}$$

Note again that the coefficients of $\mathbf{e}_x$ and $\mathbf{e}_{xy}$ exactly match the singleton potential on x and the edge potential y for the model. The residual potentials have 0 potential for the singleton x and the edge xy, and the original singleton potential for y. It is immediate that the optimal marginal vertex for the residual potentials is again v_0; if any other vertex yielded a higher score, then by ensuring that the x variable is set 0 in this new configuration, this vertex would yield a higher score than v_0 in the original model, a contradiction. Therefore we deduce that v_0 is optimal for the original model, and the inductive step is complete.

□

Lemma 8 For an arbitrary graph G, $\mathbb{L}_2(G)$ is tight for the set of balanced models.

Proof. Let $c = ((c_i)_{i\in V}, (c_{ij})_{ij\in E})$ be a balanced set of potentials on G. We may assume that all edge potentials are attractive. To see this, partition V into two disjoint sets V_a and V_b such that all edges between V_a and V_b are repulsive, whilst all within either V_a or V_b are attractive (this is possible as c is assumed balanced). Then flip the the set V_a (i.e. apply the map $F^{\dagger}_{(V_a)}$ described in §8 to c). By construction, $F^{\dagger}_{(V_a)}(c)$ has all edge potentials attractive, and by Lemma 10, $\mathbb{L}_2(G)$ is tight for c iff it is tight for $F^{\dagger}_{(V_a)}(c)$.

Let $v^\star$ be the optimal marginal vertex for c, and partition V into two disjoint sets $V_0 = \{i \in V | v^\star_i = 0\}$ and $V_1 = \{i \in V | v^\star_i = 1\}$. Given $i \in V_0$ and $j \in V_1$, note that the three tight local constraints corresponding to the edge ij are $q_{ij} \leq q_i$, $q_{ij} \geq 0$, and $q_{ij} \geq q_i + q_j - 1$. In the notation introduced in the proof of Lemma 7, the three extremal vectors contributed by this edge to the cone $N_{\mathbb{L}_2}(v^\star)$ are $\mathbf{e}_{ij} - \mathbf{e}_i$, $-\mathbf{e}_{ij}$ and $\mathbf{e}_i + \mathbf{e}_j - \mathbf{e}_{ij}$.

We construct the vector

$$\sum_{i\in V_0, j\in V_1} c_{ij}(\mathbf{e}_{ij} - \mathbf{e}_i)\,, \tag{17}$$

which lies in $N_{\mathbb{L}_2}(v^\star)$, as each vector in the sum is extremal for $N_{\mathbb{L}_2}(v^\star)$, and all coefficients are non-negative by assumption of attractiveness of c. We then consider the model given by the residual potentials c', which is the vector of potentials given by taking the difference between the original potentials and the potentials given by the vector in (17) above:

$$c'_i = c_i + \sum_{j\in V_1} c_{ij} \text{ for } i \in V_0\,, \quad c'_{ij} = 0 \text{ for } i \in V_0, j \in V_1\,, \tag{18}$$

with all other components of c' equal to the corresponding components of c . It is sufficent to show that i) the residual model has the same optimal vertex $v^\star$ as the model given by c, and ii) that the residual potentials can be written as the conical combination of extremal vectors in $N_{\mathbb{L}_2}(v^\star)$.

We begin with i). Let v' be the new optimal marginal vertex for the residual model. We further refine our partition $V = V_0 \cup V_1$ by defining

$$V_{0\to 0} = \{i \in V | v^\star_i = 0, v'_i = 0\}\,, \quad V_{0\to 1} = \{i \in V | v^\star_i = 0, v'_i = 1\}\,, \tag{19}$$
$$V_{1\to 0} = \{i \in V | v^\star_i = 1, v'_i = 0\}\,, \quad V_{1\to 1} = \{i \in V | v^\star_i = 1, v'_i = 1\}\,. \tag{20}$$

If $V_{0\to 1} = V_{1\to 0} = \emptyset$, we are done as $v^\star = v'$. Further, if $V_{0\to 1} = \emptyset$, then the score of v' for c' is the same as that for c, and it immediately follows that $v^\star$ is optimal for c' too. Also if $V_{1\to 0} = \emptyset$, it is also the case that the score of v' for c' is the same as for c, so again it follows that $v^\star$ is optimal for c'. Thus we may assume that both $V_{1\to 0}$ and $V_{0\to 1}$ are non-empty.

We introduce the following notation: $\Theta_A = \sum_{i\in A} c_i + \sum_{\substack{i,j\in A \\ \text{distinct}}} c_{ij}$, $W_{A,B} = \sum_{i\in A, j\in B \text{distinct}} c_{ij}$, for disjoint subsets $A, B \subset V$.

The score achieved by v' for the residual potentials c' is as follows, where we use the Θ and W notation as above to mean the application of the original potential c, not the new potential c':

$$\langle c', v'\rangle = \Theta_{V_{0\to 1}} + W_{V_{0\to 1}, V_1} + \Theta_{V_{1\to 1}}\,. \tag{21}$$

Since v' is optimal for c', it must have a higher score than the marginal vertex $\widetilde{v}$ given by setting all variables in the set $V_1 \cup V_{0\to 1}$ equal to 1, and all others 0. Thus, we obtain

$$\langle c', \widetilde{v}\rangle = \Theta_{V_{0\to 1}} + W_{V_{0\to 1}, V_1} + \Theta_{V_{1\to 1}} + \Theta_{V_{1\to 0}} + W_{V_{1\to 0}, V_{1\to 1}} \leq \langle c', v'\rangle\,. \tag{22}$$

By combining (21) and (22), it follows that

$$\Theta_{V_{1\to 0}} + W_{V_{1\to 0},V_{1\to 1}} \leq 0\,. \tag{23}$$

We now define the marginal vertex $\bar{v}$ as having all variables in $\mathrm{V}_{1\to 1}$ set to 1, and all others set to 0.

By examining the scores of $v^\star$ and $\bar{v}$ for the original potentials c, and recalling that $v^\star$ is optimal for c, we obtain

$$\Theta_{V_{1\to 1}} = \langle c, \bar{v}\rangle \leq \langle c, v^\star\rangle = \Theta_{V_{1\to 1}} + \Theta_{V_{1\to 0}} + W_{V_{1\to 0},V_{1\to 1}}\,. \tag{24}$$

By combining (23) and (24), we note that it must be the case that $\Theta_{V_{1\to 0}} + W_{V_{1\to 0},V_{1\to 1}} = 0$. But now we have

$$\langle c', v'\rangle = \langle c', \widetilde{v}\rangle = \langle c, \widetilde{v}\rangle \leq \langle c, v^\star\rangle = \langle c', v^\star\rangle\ , \tag{25}$$

from which it follows that $v^\star$ is optimal (over $\mathbb{M}(G)$) for c', as required for i).

It now remains to deal with point ii) - demonstrating that the residual potentials c' can be written as a conical combination of vectors in $N_{\mathbb{L}_2(G)}(v^\star)$. Consider the graph G' with vertex set V, with an edge between vertices i, j if and only if $ij \in E$ and c_{ij} is non-zero. It suffices to treat each connected component of G' separately. By verifying i) above, all edge potentials in a connected component of G' are attractive, and the optimal marginal vertex has all variables set to the same value within a connected component. Again, by flipping all variables in a connected component if necessary, we may assume that their optimal value is 0. We then demonstrate that by considering a particular vector in $N_{\mathbb{L}_2(G)}(v^\star)$, we may remove an arbitrary edge to obtain a residual model with the same optimal marginal vertex. Applying this iteratively until the residual model is a spanning tree and applying Lemma 7 then gives the result.

It therefore suffices to take an edge $ij \in E$, and consider the vector $\alpha(\mathbf{e}_{ij} - \mathbf{e}_i) + \beta(\mathbf{e}_{ij} - \mathbf{e}_j) \in N_{\mathbb{L}_2(G)}(v^\star)$, for some $\alpha, \beta \geq 0$, so that the residual model has 0 edge potential for the edge $ij \in E$, and so that the optimal marginal vertex for the model is unchanged. We now demonstrate that such α, β exist.

To ensure that the residual model has 0 edge marginal for the edge ij, we must take $\beta = W_{ij} - \alpha$. Therefore, it is no sufficient to find $0 \leq \alpha \leq W_{ij}$ so that the residual model also has optimal marginal vertex given by taking all random variables to have value 0. Denoting the potentials for the residual model by $c' = ((c'_i)_{i\in V}, (c'_{ij})_{ij\in E})$, we note that

$$c'_i = \alpha + c_i\,, \quad c'_j = (W_{ij} - \alpha) + c_j\,, \quad c'_{ij} = 0 \tag{26}$$

with all other components of c' equal to the respective component of c. Now note that any configuration of the random variables in which $X_i = X_j$ has the same score with respect to c' and with respect to c, meaning that such a configuration cannot score more highly with respect to c' than the configuration with all variables set to 0. Now consider a configuration optimal for c' subject to $X_i = 1$ and $X_j = 0$ (denote by S_i the set of indices corresponding to variables set to 1 in this configuration), and similarly consider a configuration optimal for c' subject to $X_i = 0$ and $X_j = 1$ (denote by S_j the set of indices corresponding to variables set to 1 in this configuration). It now suffices to show that both of these configurations have lower score with respect to c' than the configuration with all variables set to 0 (which has score 0). Therefore, it is sufficient to find $\alpha \geq 0$ such that

$$\alpha + \Theta_{S_i} \leq 0\,, \quad (W_{ij} - \alpha) + \Theta_{S_j} \leq 0$$

Combining these inequalities, it is sufficient to show that $W_{ij} + \Theta_{S_i} + \Theta_{S_j} \leq 0$. Note that (by optimality of the score 0 for c) we have

$$0 \geq \Theta_{S_i\cup S_j} = \Theta_{S_i} + \Theta_{S_j} - \Theta_{S_i\cap S_j} + W_{\{j\},S_i\setminus S_j} + W_{\{i\},S_j\setminus S_i} + W_{S_i\setminus S_j,S_j\setminus S_i} + W_{ij}$$

and so

$$\Theta_{S_i} + \Theta_{S_j} + W_{ij} \leq \Theta_{S_i\cap S_j} - W_{\{j\},S_i\setminus S_j} - W_{\{i\},S_j\setminus S_i} - W_{S_i\setminus S_j,S_j\setminus S_i} \leq 0$$

where the second inequality above follows as each individual term is non-positive. The result then follows as described above. □

8 Group-theoretic Results from §4.2

In this section we provide additional definitions and proofs relating to §4.2. We refer the reader to (Weller et al., 2016) for further background.

8.1 Specification of the Maps $F^{\dagger}_{(I)}$

We give a full specification of the maps $F^{\dagger}_{(I)} : \mathbb{R}^{V \cup E} \to \mathbb{R}^{V \cup E}$ for $I \subseteq V$:

$$F_{(I)}((c_i)_{i \in V}, (c_{ij})_{ij \in E})_v = -c_v - \sum_{vj \in E | v \notin I} c_{vj} \qquad v \in I\,, \tag{27}$$

$$F_{(I)}((c_i)_{i \in V}, (c_{ij})_{ij \in E})_{vw} = -c_{vw} \qquad v \in I, w \notin I\,. \tag{28}$$

with $F_{(I)}$ acting as the identity on all other coordinates. This is the natural form for such a map, as it yields the score-preserving result of Lemma 10. Note that these maps form a group under the operation of composition; in fact we have $F^{\dagger}_{(I_1)} \circ F^{\dagger}_{(I_2)} = F^{\dagger}_{(I_1 \Delta I_2)}$, where Δ is the symmetric difference operator.

8.2 Proofs of Lemmas 9 and 10

Lemma 9 $\mathbb{L}_k(G)$ is tight for a given potential $c \in \mathbb{R}^{V \cup E}$ iff it is tight for all potentials $Y^{\dagger}_{\sigma}(c)$, $\sigma \in S_V$.

Proof. It is sufficient to show that if $\mathbb{L}_k(G)$ is tight for c, then given $\sigma \in S_V$, $\mathbb{L}_k(G)$ is tight for $Y^{\dagger}_{\sigma}(c)$. Recalling that Y_{σ} maps any Sherali-Adams polytope to itself bijectively, we obtain

$$\max_{q \in \mathbb{L}_k(G)} \langle c, q \rangle = \max_{q \in \mathbb{L}_k(G)} \langle c, Y_{\sigma^{-1}} q \rangle = \max_{q \in \mathbb{L}_k(G)} \left\langle Y^{\dagger}_{\sigma}(c), q \right\rangle\,,$$

and

$$\max_{q \in \mathbb{M}(G)} \langle c, q \rangle = \max_{q \in \mathbb{M}(G)} \langle c, Y_{\sigma^{-1}} q \rangle = \max_{q \in \mathbb{M}(G)} \left\langle Y^{\dagger}_{\sigma}(c), q \right\rangle.$$

It follows that $\mathbb{L}_k(G)$ is not tight for c (i.e. $\max_{q \in \mathbb{L}_k(G)} \langle c, q \rangle > \max_{q \in \mathbb{M}(G)} \langle c, q \rangle$) if and only if $Y^{\dagger}_{\sigma}(c)$ is not tight (i.e. $\max_{q \in \mathbb{L}_k(G)} \left\langle Y^{\dagger}_{\sigma}(c), q \right\rangle > \max_{q \in \mathbb{M}(G)} \left\langle Y^{\dagger}_{\sigma}(c), q \right\rangle$).

□

Lemma 10 $\mathbb{L}_k(G)$ is tight for a given potential $c \in \mathbb{R}^{V \cup E}$ iff it is tight for all potentials $F^{\dagger}_{(I)}(c)$, $I \subseteq V$.

Proof. It is sufficient to show that if $\mathbb{L}_k(G)$ is tight for c, then given $I \subseteq V$, $\mathbb{L}_k(G)$ is tight for $F^{\dagger}_{(I)}$. Recalling that $F_{(I)}$ maps any Sherali-Adams polytope to itself bijectively, we obtain:

$$\begin{aligned}
\left\langle F^{\dagger}_{(I)}(c), F_{(I)}(q) \right\rangle =& \sum_{i \in I} F^{\dagger}_{(I)}(c)_i F_{(I)}(q)_i + \sum_{i \notin I} F^{\dagger}_{(I)}(c)_i F_{(I)}(q)_i + \\
& \sum_{ij \in E | i \in I, j \notin I} F^{\dagger}_{(I)}(c)_{ij} F_{(I)}(q)_{ij} + \sum_{ij \in E | i,j \in I \text{ or } i,j \notin I} F^{\dagger}_{(I)}(c)_{ij} F_{(I)}(q)_{ij} \\
=& \sum_{i \in I} (-c_i - \sum_{ij \in E | j \notin I} c_{ij})(1 - q_i) + \sum_{i \notin I} c_i q_i + \\
& \sum_{ij \in E | i \in I, j \notin I} -c_{ij}(q_j - q_{ij}) + \sum_{ij \in E | i,j \in I \text{ or } i,j \notin I} c_{ij} q_{ij} \\
=& \sum_{i \in I} c_i q_i + \sum_{ij \in E} c_{ij} q_{ij} - \sum_{i \in I} c_i - \sum_{ij \in E | i \in I, j \notin I} c_{ij} \\
=& \langle c, q \rangle - \sum_{i \in I} c_i - \sum_{ij \in E | i \in I, j \notin I} c_{ij}
\end{aligned}$$

Therefore, $q \in \mathbb{L}_k(G)$ yields a score for c greater than the optimum over the marginal polytope $\mathbb{M}(G)$ iff $F_{(I)}(q)$ yields a score for $F^{\dagger}_{(I)}(c)$ greater than the optimum over the marginal polytope $\mathbb{M}(G)$. From this it follows that $\mathbb{L}_k(G)$ is tight for c iff it is tight for $F^{\dagger}_{(I)}(c)$. □

9 Proof of the Marginals Extension Result for $\mathbb{L}_3$

In this section, we provide a proof of the following result.

Theorem 4 If G is a chordal graph, then the polytope given by enforcing TRI constraints on just the triplets of distinct variables that form 3-cliques in G is equal to $\mathbb{L}_3(G)$, the polytope given by enforcing constraints on all triplets of distinct variables in G.

An equivalent definition of chordality is the following.

Definition 13. A graph $G = (V, E)$ is chordal if there exists a *perfect elimination ordering* for G. That is, if there exists an ordering $v_1, \ldots, v_n$ of the vertex set V such that the set $v_i \cup (\mathrm{N}_G(v_i) \cap \{v_{i+1}, \ldots, v_n\})$ is a clique for all $i = 1, \ldots, n$, where $\mathrm{N}_G(v) = \{w \in V \,|\, vw \in E\}$ is the neighbor set for vertex $v \in V$.

From this definition, it is straightforward to show the following result, which shows that edges may be added sequentially to a chordal graph to make it complete without breaking the property of chordality at any point.

Lemma 14. *If $G = (V, E)$ is a non-complete chordal graph, then there exist distinct $i, j \in V$ such that $ij \notin E$, and the graph with vertex set V and edge set $E \cup \{i, j\}$ is chordal.*

Proof. Let $v_1, \ldots, v_n$ be a perfect elimination ordering of the vertex set V for G. Let m be the largest element of $\{1, \ldots, n\}$ such that the set $\{v_m, \ldots, v_n\}$ is not a clique. Such an m exists as G is assumed non-complete. We may assume without loss of generality that $\{v_m, v_{m+1}\} \notin E$: there exists some vertex $v \in \{v_{m+1}, \ldots, v_n\}$ such that $\{v_m, v\} \notin E$ (else $\{v_m, \ldots, v_n\}$ would be a clique), and swapping v with v_{m+1} leads to another valid perfect elimination ordering with the desired property that $\{v_m, v_{m+1}\} \notin E$.

We now demonstrate that $v_1, \ldots, v_n$ is a perfect elimination ordering for the graph G' with vertex set V and edge set $E \cup \{v_i, v_{i+1}\}$, which is the original graph G with an added edge $\{v_m, v_{m+1}\}$, is chordal. For $j > m$, note that we have $G[v_j \cup (\mathrm{N}_G(v_j) \cap \{v_{j+1}, \ldots, v_n\})] = G'[v_j \cup (\mathrm{N}_{G'}(v_j) \cap \{v_{j+1}, \ldots, v_n\})]$, and this subgraph is therefore complete, meaning that $v_j \cup (\mathrm{N}_{G'}(v_j) \cap \{v_{j+1}, \ldots, v_n\})$ is a clique in G'. For $j < m$, note that $\mathrm{N}_G(v_j) = \mathrm{N}_{G'}(v_j)$, but since $G[v_j \cup (\mathrm{N}_G(v_j) \cap \{v_{j+1}, \ldots, v_n\})]$ is complete, it follows that $G'[v_j \cup (\mathrm{N}_{G'}(v_j) \cap \{v_{j+1}, \ldots, v_n\})]$ is complete too, and hence $v_j \cup (\mathrm{N}_{G'}(v_j) \cap \{v_{j+1}, \ldots, v_n\})$ is a clique in G'. Finally we check that $v_m \cup (\mathrm{N}_{G'}(v_m) \cap \{v_{m+1}, \ldots, v_n\})$ is a clique in G'. By construction of m, the set $\{v_{m+1}, \ldots, v_n\}$ forms a clique in G and hence in G'. Therefore $v_m \cup (\mathrm{N}_{G'}(v_m) \cap \{v_{m+1}, \ldots, v_n\})$ is a clique, as it is formed of the clique $\mathrm{N}_{G'}(v_m) \cap \{v_{m+1}, \ldots, v_n\}$ with the addition of the vertex v_{m+1}, which has been shown to have an edge to every vertex in the set $\{v_m\} \cup \{v_{m+2}, \ldots, v\}$. □

We now set out a key proposition which we will be able to apply iteratively in combination with Lemma 14 to prove Theorem 4.

Proposition 15. If $G = (V, E)$ is a non-complete chordal graph, then given a configuration $((q_i)_{i \in V}, (q_{ij})_{ij \in E})$ satisfying all TRI constraints involving $3-$cliques in G, there exists an edge e not present in E, and a corresponding pseudomarginal q_e such that: i) $G' = (V, E \cup \{e\})$ is chordal; ii) the augmented configuration $((q_i)_{i \in V}, (q_{ij})_{ij \in E})$ satisfies all triplet constraints enforced on 3-cliques of G'.

Proof. i) This is the statement of Lemma 14.

ii) Lemma 14 can be used to select the additional edge ij to be added to E. We consider the new TRI constraints that will be enforced by adding this edge; they are exactly those corresponding to triplets of variables made up of i, j, and one other variable, say k, such that $ik \in E, jk \in E$. For such a triplet, the TRI constraints (see §10.4.1) enforce the following inequalities on the new pseudomarginal q_{ij}:

$$\max(q_{ik} + q_{jk} - q_k, q_i + q_j + q_k - q_{ik} - q_{jk} - 1) \leq q_{ij} \leq \min(q_i + q_{jk} - q_{ik}, q_j + q_{ik} - q_{jk}) \tag{29}$$

Therefore each $k \in K \coloneqq \{v \in V | iv, jv \in E\}$ enforces an interval I_k of settings for q_{ij} which are valid for the TRI constraints over the triplet i, j, k. Showing that there is an assignment to q_{ij} that respects all TRI constraints simultaneously is equivalent to showing that $\cap_{k \in K} I_k$ is non-empty. Since all the I_k are intervals, it suffices to show that their pairwise intersections are all non-empty. To see this, note that $\cap_{k \in K} I_k = [\max_k \inf I_k, \min \sup I_k]$, so it is enough that for each $k, l \in K$, we have $\inf I_k \leq \sup I_l$. Therefore, it is sufficient to show that for any $k_1, k_2 \in K$, the right-hand side of the inequalities of the form (29) induced by k_1 is greater than or equal to the left-hand side of the inequalities of the form (29)

induced by k_2. In particular, it suffices to demonstrate the following four inequalities

$$q_{ik_2} + q_{jk_2} - q_{k_2} \leq q_i + q_{jk_1} - q_{ik_1} \,, \tag{30}$$
$$q_{ik_2} + q_{jk_2} - q_{k_2} \leq q_j + q_{ik_1} - q_{jk_1} \,, \tag{31}$$
$$q_i + q_j + q_{k_2} - q_{ik_2} - q_{jk_2} - 1 \leq q_i + q_{jk_1} - q_{ik_1} \,, \tag{32}$$
$$q_i + q_j + q_{k_2} - q_{ik_2} - q_{jk_2} - 1 \leq q_j + q_{ik_1} - q_{jk_1} \,. \tag{33}$$

By symmetry in i and j, it is enough to show Inequalities (30) and (32), since Inequalities (31) and (33) are given by permuting indices i and j in (30) and (32). Note that since we have $\{i, k_1\}, \{j, k_1\}, \{i, k_2\}, \{j, k_2\} \in E$ and $\{i, j\} \notin E$, chordality of G implies that $\{k_1, k_2\} \in E$. In particular, it follows that all TRI constraints for the triplets i, k_1, k_2 and j, k_1, k_2 are satisfied. By adding the inequalities $q_i + q_{k_1 k_2} \geq q_{ik_1} + q_{ik_2}$ and $q_{k_2} + q_{jk_1} \geq q_{jk_2} + q_{k_1 k_2}$, we obtain Inequality (30). By adding the TRI inequalities $q_{jk_1} + q_{jk_2} + q_{k_1 k_2} \geq q_j + q_{k_1} + q_{k_2} - 1$ and $q_{k_1} + q_{ik_2} \geq q_{k_1 k_2} + q_{ik_1}$, we obtain Inequality (32). □

A proof of Theorem 4 is now straightforward.

Proof of Theorem 4. We may apply Proposition 15 inductively, to iteratively yield an edge not already in the graph, and a valid pseudomarginal for that edge, until we have a complete graph. □

10 Proof of Theorem 3 for $\mathbb{L}_3^s$

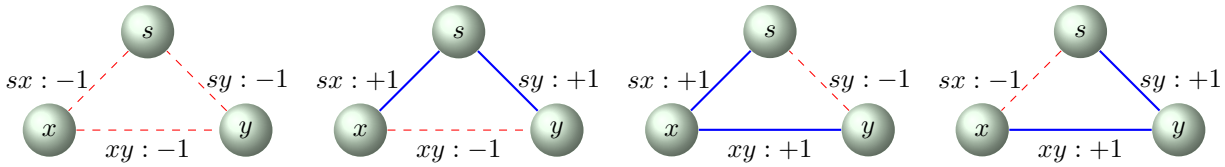
In this section, we provide a proof of the following result.

Theorem 3 Given a graph $G = (V, E)$, the polytope $\mathbb{L}_3^s(G)$ is tight on the class of almost balanced models with distinguished variable s. Further, there are no linear constraints of $\mathbb{L}_3^s(G)$ that may be removed and yield a polytope which is tight on this class of potentials.

10.1 Proof of the Second Claim of Theorem 3

We first give a short proof for the second claim of Theorem 3, and then discuss the structure of the proof for the first claim.

Consider the polytope given by enforcing all linear constraints of $\mathbb{L}_3^s(G)$ except one TRI constraint, say on a triangle sxy. There are four possible inequalities to remove: a) $q_{sx} + q_{sy} + q_{xy} \geq q_s + q_x + q_y - 1$; b) $q_s + q_{xy} \geq q_{sx} + q_{sy}$; c) $q_x + q_{sy} \geq q_{sx} + q_{xy}$; d) $q_y + q_{sx} \geq q_{sy} + q_{xy}$. In each case, we consider a model with all potentials apart from those corresponding to the singletons s, x and y and the edges sx, sy, xy set to 0. We set the singleton potentials $\theta_s, \theta_x, \theta_y$ and edge potentials W_{sx}, W_{sy}, W_{xy} as in the corresponding Figures 9a to 9d. We now show that these potentials are non-tight.



(a) $\theta_s = \theta_x = \theta_y = 1$ (b) $\theta_s = \theta_x = \theta_y = -1/2$ (c) $\theta_s = \theta_x = \theta_y = -1/2$ (d) $\theta_s = \theta_x = \theta_y = -1/2$

Figure 9: Solid blue edges are attractive, dashed red edges are repulsive.

- *Inequality a).* In this case the MAP score is $1/2$, attained by the configuration $(X_s, X_x, X_y) = (1, 0, 0)$. A valid configuration in the polytope given by $\mathbb{L}_3^s(G)$ without inequality a) is $(q_s, q_x, q_y, q_{sx}, q_{sy}, q_{xy}) = (1/2, 1/2, 1/2, 0, 0, 0)$, which has score $3/4$, demonstrating non-tightness of this polytope.
- *Inequality b).* In this case the MAP score is 0, attained by the configuration $(X_s, X_x, X_y) = (0, 0, 0)$. A valid configuration in the polytope given by $\mathbb{L}_3^s(G)$ without inequality b) is $(q_s, q_x, q_y, q_{sx}, q_{sy}, q_{xy}) = (1/2, 1/2, 1/2, 1/2, 1/2, 0)$, which has score $1/4$, demonstrating non-tightness of this polytope.
- *Inequality c).* In this case the MAP score is 0, attained by the configuration $(X_s, X_x, X_y) = (0, 0, 0)$. A valid configuration in the polytope given by $\mathbb{L}_3^s(G)$ without inequality b) is $(q_s, q_x, q_y, q_{sx}, q_{sy}, q_{xy}) = (1/2, 1/2, 1/2, 1/2, 0, 1/2)$, which has score $1/4$, demonstrating non-tightness of this polytope.

- *Inequality d).* In this case the MAP score is 0, attained by the configuration $(X_s, X_x, X_y) = (0, 0, 0)$. A valid configuration in the polytope given by $\mathbb{L}_3^s(G)$ without inequality b) is $(q_s, q_x, q_y, q_{sx}, q_{sy}, q_{xy}) = (1/2, 1/2, 1/2, 0, 1/2, 1/2)$, which has score $1/4$, demonstrating non-tightness of this polytope.

10.2 Structure of the Proof for the First Claim of Theorem 3

The structure of the proof of the first claim of Theorem 3 broadly follows that of the proof of Theorem 2, which appears in (Weller et al., 2016). First note that it is sufficient to prove the result for almost attractive models, since given an almost balanced model, there exists an element of the flipping group $\mathbb{Z}_2^{|V|}$ (see §4.2) that renders the model almost attractive, and preserves properties of (non-)tightness. Let $c = ((\theta_v)_{v\in V}, (W_e)_{e\in E})$ be a set of potentials corresponding to an almost balanced graphical model with distinguished variable s, so that the probability of a given configuration $v \in \mathrm{Ext}(\mathbb{M}(G))$ is proportional to $\exp(\langle c, v\rangle)$. Define the function $F^s_{\mathbb{L}_3^s(G)} : [0, 1] \to \mathbb{R}$ by

$$F^s_{\mathbb{L}_3^s(G)}(t) = \sup_{q\in\mathbb{L}_3^s(G)\big|_{q_s=t}} \langle c, q\rangle$$

We will show that $F^s_{\mathbb{L}_3^s(G)}$ is linear, and so its maximal value is attained for $t = 0$ or $t = 1$. This shows that a global maximizer of θ is achieved over $\mathbb{L}_3^s(G)$ when $q_s = 0$ or $q_s = 1$. But we know that $\mathbb{L}_3^s\big|_{q_s=0}$ (resp. $\mathbb{L}_3^s\big|_{q_s=1}$) is tight wrt $\mathbb{M}\big|_{v_s=0}$ (resp. $\mathbb{M}_{v_s=1}$), since these problems are equivalent to MAP inference problems over $G \setminus \{s\}$, which is balanced, and regarding the polytopes $\mathbb{L}_3^s\big|_{q_s=0}$ and $\mathbb{L}_3^s\big|_{q_s=1}$ as polytopes over this reduced model, they are both equal to $\mathbb{L}_2(G\setminus\{s\})$, which is tight on balanced models. The result then follows.

10.3 Concavity

Demonstrating the concavity of $F^s_{\mathbb{L}_3^s(G)}$ is straightforward; the proof below is an adaptation of (Weller et al., 2016, Lemma 5).

Lemma 16. *$F^s_{\mathbb{L}_3^s(G)}$ is concave.*

Proof. Let $t_1, t_2 \in [0, 1]$, and let $\lambda \in [0, 1]$. Let $q_i^\star \in \arg\max_{q\in\mathbb{L}_3^s|_{q_s=t_i}} \langle\theta, q\rangle$ for $i = 1, 2$, so that we have

$$F^s_{\mathbb{L}_3^s(G)}(t_i) = \langle c, q_i^\star\rangle$$

for $i = 1, 2$. Let $\widetilde{t} = \lambda t_1 + (1 - \lambda)t_2$. Then consider $\widetilde{q} = \lambda q_1^\star + (1 - \lambda)q_2^\star$. By convexity, $\widetilde{q} \in \mathbb{L}_3^s|_{q_s=\widetilde{t}}$. By linearity of the score, we have

$$F^s_{\mathbb{L}_3^s(G)}(\widetilde{t}) \geq \langle c, \widetilde{q}\rangle = \lambda \langle c, q_1^\star\rangle + (1 - \lambda) \langle c, q_2^\star\rangle = \lambda F^s_{\mathbb{L}_3^s(G)}(t_1) + (1 - \lambda)F^s_{\mathbb{L}_3^s(G)}(t_2)$$

which demonstrates concavity. □

10.4 Convexity

We now prove that $F^s_{\mathbb{L}_3^s(G)}$ is convex. To do this, we show that elements of the set $\arg\max_{q\in\mathbb{L}_3^s|_{q_s=x}} \langle\theta, q\rangle$ must have very specific structural properties. We begin by introducing the structural properties that will be relevant.

10.4.1 Local and Triplet Inequalities

We describe the explicit form of the local inequalities (the pairwise inequalities each involving two variables sharing an edge, describing the polytope $\mathbb{L}_2(G)$) and the triplet inequalities (which together with the local inequalities describe the polytope $\mathbb{L}_3(G)$).

From the definition of the polytope $\mathbb{L}_2(G)$ (see Definition 1), we need to ensure that for each pair of distinct vertices $i, j \in V$, there exists a valid joint distribution over the variables X_i and X_j. Recalling that q_{ij} represents the probability $\mathbb{P}(X_i = 1, X_j = 1)$ and q_i (respectively q_j) represent $\mathbb{P}(X_i = 1)$ (respectively $\mathbb{P}(X_j = 1)$), we have the following

parametrization for the atomic events associated with X_i and X_j:

$$\begin{aligned}
&\mathbb{P}(X_i = 1, X_j = 1) = q_{ij}\,, \\
&\mathbb{P}(X_i = 0, X_j = 1) = q_j - q_{ij}\,, \\
&\mathbb{P}(X_i = 1, X_j = 0) = q_i - q_{ij}\,, \\
&\mathbb{P}(X_i = 1, X_j = 1) = 1 - q_i - q_j + q_{ij}\,.
\end{aligned}$$

By the form of our parametrization these quantities sum to 1, so it is sufficient to enforce that they are all non-negative to ensure that the parameters q_i, q_j, and q_{ij} correspond to a valid distribution. Enforcing these non-negativity constraints yields the linear constraints of the polytope $\mathbb{L}_2(G)$:

$$\max(0, q_i + q_j - 1) \leq q_{ij} \leq \min(q_i, q_j) \qquad \forall i, j \in V\,.n \tag{34}$$

The additional constraints needed to describe the polytope $\mathbb{L}_3(G)$ can be derived in the same way; we now need to ensure that for every triplet of distinct indices $i, j, k \in V$, there exists a joint distribution over X_i, X_j, X_k such that $q_i, q_j, q_k, q_{ij}, q_{ik}, q_{jk}$ are marginals of this distribution. The atomic events associated with this distribution over 3 variables can be expressed in terms of these 6 parameters and one additional parameter α, which will represent the probability of all three variables taking the value 1. We then have

$$\begin{aligned}
&\mathbb{P}(X_i = 1, X_j = 1, X_k = 1) = \alpha\,, && \mathbb{P}(X_i = 1, X_j = 0, X_k = 0) = q_i - q_{ij} - q_{ik} + \alpha \\
&\mathbb{P}(X_i = 0, X_j = 1, X_k = 1) = q_{jk} - \alpha\,, && \mathbb{P}(X_i = 0, X_j = 1, X_k = 0) = q_j - q_{ij} - q_{jk} + \alpha \\
&\mathbb{P}(X_i = 1, X_j = 0, X_k = 1) = q_{ik} - \alpha\,, && \mathbb{P}(X_i = 0, X_j = 0, X_k = 1) = q_k - q_{ik} - q_{jk} + \alpha \\
&\mathbb{P}(X_i = 1, X_j = 1, X_k = 0) = q_{ij} - \alpha\,, && \mathbb{P}(X_i = 0, X_j = 0, X_k = 0) = 1 - q_i - q_j - q_k + q_{ij} + q_{ik} + q_{jk} - \alpha
\end{aligned}$$

Again, by construction of the parametrization these eight quantities sum to 1, so we just need enforce non-negativity of each entry to ensure that they describe a valid joint distribution. This gives rise to eight inequalities; performing Fourier-Motzkin elimination of the variable α in these inequalities then yields the local inequalities derived above for each pair of vertices (i, j and i, k and j, k), as well as four additional triplet inequalities involving all three vertices:

$$\begin{aligned}
&q_i + q_{jk} \geq q_{ij} + q_{ik} \\
&q_j + q_{ik} \geq q_{ij} + q_{jk} \\
&q_k + q_{ij} \geq q_{ik} + q_{jk} \\
&q_{ij} + q_{ik} + q_{jk} \geq q_i + q_j + q_k - 1 \qquad (35) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (36)
\end{aligned}$$

We refer to these inequalities as the triplet (or TRI) inequalities. The polytope $\mathbb{L}_3(G)$ is obtained by enforcing the local constraints (34) for each pair of distinct vertices in V, and the TRI inequalities (35) for each triplet of distinct vertices in V. To obtain the polytope $\mathbb{L}_3^s(G)$ (see §3 for definition), we enforce the local constraints for each pair of distinct vertices of vertices in V, and the TRI inequalities for each triplet of distinct vertices in V which includesp the variable s.

10.4.2 Strong and Locking Edges

Given $q = ((q_i)_{i \in V}, (q_{ij})_{ij \in E}) \in \mathbb{L}_3^s(G)$, we say that edge ij is strong if a local constraint is tight for vertices i and j , so that $q_{ij} = \min(q_i, q_j)$ or $q_{ij} = \max(0, q_i + q_j - 1)$; in the case of the former, we say that edge ij is *strong up*, and in the case of the latter, *strong down*. We say that a cycle $i_1 i_2, \ldots, i_m i_1$ in G is strong frustrated if all edges are strong, and there are an odd number of strong down edges. We say that variables i and j are *locking up* (or *locked up*) if $q_i = q_j = q_{ij}$, and that variables i and j are *locking down* (or *locked down*) if $q_i = 1 - q_j$, and $q_{ij} = 0$.

10.4.3 Main Structural Result for Optimal Vertices of the Polytope $\mathbb{L}_3^s|_{q_s = x}$

With these definitions in hand, we will demonstrate the following:

Theorem 17. *Any element of* $\arg\max_{q \in \mathbb{L}_3^s|_{q_s = x}} \langle c, q \rangle$ *must have all singleton marginals in the set* $\{0, x, 1 - x, 1\}$*, any edges between variables with singleton marginals in the set* $\{x, 1 - x\}$ *must be locking, and there are no strong frustrated cycles in the configuration.*

With Theorem 17, we can then deduce convexity of $F^s_{\mathbb{L}^s_3(G)}$ immediately by using the argument in (Weller et al., 2016, §14). Namely, we may write

$$F^s_{\mathbb{L}^s_3(G)}(t) = \sup_{y\in[0,1]} \langle c, q(t|y)\rangle \ , \tag{37}$$

where $q(\cdot|y) : [0,1] \to \mathbb{L}^s_3(G)$ is defined as follows. Let $q^\star$ be optimal for c in $\mathbb{L}^s_3|_{q_s=y}$. By Theorem 17, the sets $A_y = \{j : q^\star_j = 0\}, B_y = \{j : q^\star_j = y\}, C_y = \{j : q^\star_j = 1-y\}$ and $D_y = \{j : q^\star_j = 1\}$ exhaustively partition the set V. Now define the components of $q(\cdot|y)$ as follows:

$$q_j(t|y) = \begin{cases} 0 & j \in A_y \\ t & j \in B_y \\ 1-t & j \in C_y \\ 1 & j \in D_y \end{cases}, \qquad q_{ij}(t|y) = \begin{cases} 0 & i \in A_y \text{ or } j \in A_y \\ q_i & j \in D_y \\ q_j & i \in D_j \\ t & i,j \in B_y \\ 1-t & i,j \in C_y \\ 0 & i \in B_y \text{ and } j \in C_y; \text{ or } i \in C_y \text{ and } j \in B_y. \end{cases}$$

Note that $q(t|y) \in \mathbb{L}^s_3(G)$ for all $t \in [0,1]$ since all edges of $q(t|y)$ are locking and there are no frustrated cycles. From Equation (37), note that $F^s_{\mathbb{L}^s_3(G)}$ is a supremum of linear functions, and is therefore convex.

We thus turn our attention to proving Theorem 17.

10.5 Structure and Locking Edges

We work up to proving Theorem 17 via several intermediate results. We first prove a structural result on the optimal vertices of the polytope $\mathbb{L}^s_3|_{q_s=x}$ for the potential c, assuming no locking edges or singletons equal to 0 or 1 at the extremal vertex. These assumptions will be relaxed to obtain a more general result, Proposition 23, by using an argument that shows if an optimal vertex has locking edges, these can be safely "contracted away" for the sake of analysis, and the results proven for reduced models with no locking edges apply, which may then be "expanded" back up to the full model.

10.5.1 Structure for Models with No Locking Edges at Optimal Vertices

We state and prove two key structural results for models such that optimal vertices contain no locking edges.

Proposition 18. Suppose $q \in \arg\max_{q\in\mathbb{L}^s_3|_{q_s=x}} \langle c, q\rangle$ is extremal, and moreover suppose there are no locking edges in q, and no singleton marginals equal to 0 or 1. Then all edges involving s are weak.

Proof. Suppose the conclusion does not hold; let q be such that not all edges involving s are weak. Since we assume no locking edges, any variable $v \in V \setminus \{s\}$ such that sv is strong cannot have marginal $q_v \in \{x, 1-x\}$. Then by sections 12 and 13 of Weller et al. (2016), there exists a non-zero perturbation $p \in \mathbb{R}^{V\cup E}$ (as there is some strong edge involving s), and is such that any constraints of the polytope $\mathbb{L}_3(G)$ which are tight at q remain tight at $q \pm p$. Since $\mathbb{L}^s_3(G)$ is formed of a subset of the constraints of $\mathbb{L}_3(G)$, it follows that $q+p, q-p \in \mathbb{L}^s_3(G)$. Further, by construction the perturbation does not move q_s; q is therefore a convex combination of two other points in $\mathbb{L}^s_3|_{q_s=x}$, and is therefore not extremal in this polytope. □

Proposition 19. Suppose $q \in \arg\max_{q\in\mathbb{L}^s_3|_{q_s=x}} \langle c, q\rangle$ is extremal, and moreover suppose there are no locking edges in q, and no singleton marginals equal to 0 or 1. Then it can only be the case that the graph underlying the model consists of the single vertex s.

Proof. Suppose not. By Proposition 18, all edges of q involving s are weak. There are therefore no strong down edges in q; as all edge potentials away from s are attractive, if an edge ab away from s is strong down, it must be because it is being held down by a TRI constraint (involving s); without loss of generality, we have $q_a + q_{sb} = q_{ab} + q_{as}$. If q_{as} is strong down and equal to 0, then we have $q_a = q_{as} - q_{sb} \geq q_{as} \geq q_a$, and so we deduce that $q_{as} = q_a$ and $q_{sb} = 0$, contradicting the fact that there are no strong down edges in q. A contradiction may similarly be reached in the case where q_{as} is strong down and equal to $q_a + q_s - 1$.

Therefore, as argued in Section 14 of Weller et al. (2016), the following perturbation p is such that all TRI constraints which are tight at q remain tight at $q \pm \varepsilon p$, for sufficiently small $\varepsilon > 0$ and therefore $q \pm \varepsilon p \in \mathbb{L}^s_3(G)$, yielding a contradiction unless the underlying model is a single variable s.

$$p_v = \begin{cases} p_s = 0 & v = s \\ p_v = +1 & v \in V \setminus \{s\} \end{cases}, \quad p_{uv} = \begin{cases} +\frac{1}{2} & uv \text{ edge, with } u = s \\ +1 & uv \text{ edge, with } u, v \in V \setminus \{s\} \end{cases}.$$

□

10.5.2 Structural Results Concerning Locking Edges

With these results in hand, we now relax the assumptions of having no locking edges in the arg max configuration q. To do this, we first demonstrate that the relation on the set of variables defined by $i \sim j$ if the edge ij is locking (and prescribing that $i \sim i \; \forall i \in V$) is an equivalence relation, and use this to show that these *locking components* can, in some sense, be shrunk down to a single variable for the purposes of understanding the behavior of the arg max. Reflexivity of $\sim$ is included in the definition, and symmetry is immediate. It remains to show transitivity. We provide a full proof of this result as Lemma 22, first giving intermediate results in Lemmas 20 and 21.

Our first result deals specifically with locked up edges.

Lemma 20. *Locking up is a transitive relation on the set of vertices of the graph.*

Proof. Suppose $a \sim b$ and $b \sim c$. Consider the triangle sab. TRI inequalities imply that $q_{sa} = q_{sb}$. Similarly, TRI inequalities in sbc imply that $q_{sb} = q_{sc}$. Note that ac is locked up iff $q_{ac} = q_a$. Suppose this does not hold. Since the potential associated to edge ac is attractive, either sa or sc is holding ac down in a TRI constraint. By symmetry, suppose sa is holding down ac. Then

$$\begin{aligned} & q_c + q_{sa} = q_{sc} + q_{ac} \\ \implies & q_a + q_{sa} = q_{sa} + q_{ac} \\ \implies & q_a = q_{ac} \end{aligned}$$

and so we deduce that $a \sim c$. □

We then use the following lemma to deal with locking down edges.

Lemma 21. *i) If two variables a, b are locked down, then the edges sa and sb are also locked. ii) For any variables x, y such that sx and sy are both locked, then xy is locked too.*

Proof. For i), let $q_b = 1 - q_a$, and $q_{ab} = 0$. Since ab is attractive, this edge is being held down by some TRI inequality in the triangle sab. Without loss of generality, $q_a + q_{sb} = q_{sa} + q_{ab}$, so $q_a + q_{sb} = q_{sa}$, so $q_a = q_{sa}$ and $q_{sb} = 0$. Now consider the following TRI inequality:

$$\begin{aligned} & q_{sa} + q_{sb} + q_{ab} \geq q_s + q_a + q_b - 1 \\ \implies & q_a + 0 + 0 \geq q_s + q_a + (1 - q_a) - 1 \\ \implies & q_a \geq q_s \end{aligned}$$

But $q_{sa} = q_a$, so $q_a \leq q_s$, so $q_s = q_a$, and so sa is locked up, and sb is locked down.

For ii), we show that any two variables locking to s lock with one another. Suppose x, y are locked to s. By flipping s if necessary, we need not deal with the case where x and y are both locked up with s.

Consider the case where sx is locked up, and sy is locked down. Then consider the TRI inequality $q_x + q_{sy} \geq q_{sx} + q_{xy}$. Substituting our values in , this gives $q_x + 0 \geq q_x + q_{xy}$. So $q_{xy} = 0$, and so q_{xy} is locking down.

Now consider the case where sx and sy are locked down. Consider the following TRI inequality

$$\begin{aligned} & q_{sx} + q_{sy} + q_{xy} \geq q_s + q_x + q_y - 1 \\ \implies & q_{xy} \geq s + (1 - q_s) + (1 - q_s) - 1 \\ \implies & q_{xy} \geq 1 - q_s \end{aligned}$$

But $q_{xy} \leq \min(q_x, q_y) = 1 - q_s$, so $q_{xy} = 1 - q_s$ and xy is locked up. □

Finally, the results above are combined to provide a full proof of transitivity.

Lemma 22. *Locking is an equivalence relation on the set of vertices of the graph.*

Proof. Let $a \sim b$, $b \sim c$; we aim to show that $a \sim c$. If ab is locked up and bc is locked up, then Lemma20 applies immediately to give $a \sim c$. If both ab and ac are locked down, then we have from Lemma 21 i) that sa, sb, sc are all locked, and by applying Lemma 21 ii) to sa and sc, we deduce that $a \sim c$. Finally, if ab is locked down and bc is locked up, Lemma 21 i) applies to give either that sa is locked down and sb is locked up, or that sa is locked up and sb is locked down. In the case of the former, note that sb and bc are locked up, so sc is locked up by Lemma 20, and so ac is locked up by Lemma 21 ii). In the case of the latter, we consider the TRI inequality $q_c + q_{sb} \geq q_{sc} + q_{bc}$, and deduce that sc is locked, and so by Lemma 21 ii), ac is locked. This completes the proof. □

10.5.3 Proof of Main Structural Result

The results of §10.5.1 and §10.5.2 are now combined to yield the main structural result needed to deduce Theorem 17

Proposition 23. Suppose $q \in \arg\max_{q \in \mathbb{L}_3^s|_{q_s = x}} \langle \theta, q \rangle$ is extremal. Then all singleton marginals lie in the set $\{0, x, 1 - x, 1\}$, and all variables lie in a single locking component.

Proof. We first may eliminate any variables with $0, 1$ singleton marginals. We then consider the equivalence classes of variables induced by the equivalence relation of having a locking edge, and then consider the submodel induced by selecting one representative from each equivalence class. We apply Propositions 18 and 19 to obtain a perturbation for this reduced model, and show that this perturbation can be "expanded" to the original model, so that the conclusions of the proposition hold. The perturbation is expanded by specifying that if a variable a is locked up to its representative, it moves in the same direction as the representative, whilst if it is locked down, it moves in the opposite direction to its representative. Note that as all locking components contain no frustrated cycles, it follows that if two variables are locked up they are perturbed in the same direction, whilst if they are locked down, they move in opposite directions. Strong edges are re-optimized, weak edges do not move. It will then follow that the conclusion of Proposition 19 holds for the reduced model, and so the expanded model must have a single locking component, and the conclusion of the proposition follows.

To be precise, we need to verify the following:

1. If the TRI inequalities corresponding to a triangle sab are satisfied under the perturbation, and b is locked to b', then the TRI inequalities associated with the triangle sab' are also satisfied under the perturbation.

2. If the LOC constraints corresponding to a and b are satisfied under the perturbation, and b' is locked (either up or down) to b, then the LOC constraints corresponding to a and b' are satisfied under the perturbation.

We deal with these in turn:

Firstly, let s, a, b, b' be as mentioned in i) above, and assume b and b' are locked up - see Figure 10.

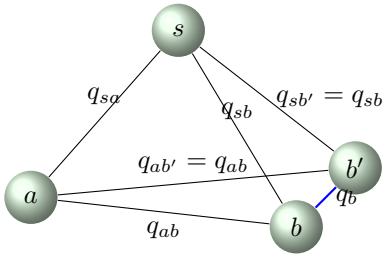


Figure 10: One case to consider from point 1. above.

Consider the triangle sbb'. By considering the inequalities $q_b + q_{sb'} \geq q_{sb} + q_{bb'}$ and $q_{b'} + q_{sb} \geq q_{sb'} + q_{bb'}$, we deduce that $q_{sb} = q_{sb'}$. We can also deduce that $q_{ab} = q_{ab'}$; if they are not both strong up (which would allow us to deduce that they are equal), then one of them is held down by a TRI constraint. However, since all singleton and edge marginals in triangles sab and sab' match (apart from maybe q_{ab} and $q_{ab'}$), they must be held to the same value by their respective TRI

constraints, and are hence equal. Therefore, since b' is specified to move with b, the edges sb' and ab' move with sb and ab respectively, and so any tight TRI constraint in sab' remains tight under the perturbation.

Now consider the case where b and b' are locked down. By Lemma 21, b and b' are both locked to s - one is locked up, one locked down. By flipping s, we may assume that sb is locked down, and sb' is locked up. We may assume a is not locking with s or b - if it were, then all singleton and edge marginals would be fixed under the perturbation, and so any TRI inequalities would trivially be preserved.

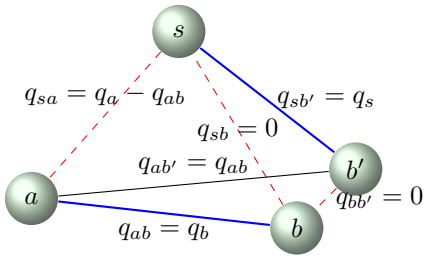


Figure 11: Solid blue edges are attractive, dashed red edges are repulsive.

Now sab has a single locking edge, from this it follows that there are two tight TRI constraints, namely $q_a + q_{sb} = q_{sa} + q_{ab}$, and $q_{sa} + q_{sb} + q_{ab} = q_s + q_a + q_b - 1$. This leads to $q_{sa} = q_a - q_{ab}$. Considering sab', again two TRI constraints must be tight, namely $q_a + q_{sb'} = q_{sa} + ab'$ and $q_{sa} + q'_{sb} + q'_{ab} = q_s + q_a + q'_b - 1$, which leads to $q'_{ab} = q_{sa}$. Now note that the perturbation moves exactly the same set of singleton and edge marginals in sab as it does in sab' (b, b' are both locked to s so don't move), and the same TRI constraints are tight in each triangle. Therefore, the fact that all TRI constraints in sab are satisfied under the perturbation implies the same is true for sab', as required.

Now consider a, b, b' as mentioned in ii) above. The case where one of these variables is s is dealt with the argument for i). When none of the variables are s, this point is in fact also dealt with by the argument for i) - we have shown above that if the perturbation respects TRI constraints for sab, then it respects TRI constraints for sab'. But TRI constraints being respected for sab' implies that LOC constraints are respected for ab', so we are done. □

Theorem 17 then follows from Proposition 23, from which in turn the linearity of $F^s_{\mathbb{L}^s_3(G)}$ follows from the discussion in §10.4.3, and hence Theorem 3 follows.

11 Identifying Non-tight Signings of T_4, Treewidth 3 Forbidden Minors

We provide a proof of the following result from §5.

Theorem 12 The only non-tight signing for $\mathbb{L}_4$ of any minimal forbidden minor for treewidth ≤ 3 is the odd-K_5.

Recall from §5 that for a given signed graph (G, Σ), $\mathbb{L}_4(G)$ is tight for all signings of G with signed topology Σ iff the following problem has optimal value 0:

$$\max_{c \in \widetilde{\Sigma}} \max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} [\langle c, q \rangle - \langle c, x \rangle] \tag{38}$$

11.1 Linearizing the Problem

We first convert this indefinite quadratic program over a conic region to an optimization problem over a polytope, by intersecting the feasible region with some convex polytope containing the origin in its interior:

Lemma 24. *Problem* (38) *has a non-zero optimum iff*

$$\max_{c \in \widetilde{\Sigma} \cap P} \max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} [\langle c, q \rangle - \langle c, x \rangle] \tag{39}$$

has a non-zero optimum, where P is any polytope containing the origin in its interior.

Proof. Since $\widetilde{\Sigma} \cap P \subseteq \widetilde{\Sigma}$, if (39) has a non-zero optimum then so does (38). For the opposite direction, suppose (38) has a non-zero optimum, so in particular there exist $c' \in \mathbb{R}^{V \cup E}$, $q' \in \mathbb{L}_4(G)$ and $x' \in \mathbb{M}(G)$ such that

$$\langle c', q' \rangle - \langle c', x' \rangle > 0$$

As $\mathbf{0} \in \operatorname{int}(P)$, P contains an open ball around the origin, and so there exists $\lambda > 0$ such that $\lambda c' \in P$. Then note

$$\langle \lambda c', q' \rangle - \langle \lambda c', x' \rangle = \lambda \left(\langle c', q' \rangle - \langle c', x' \rangle \right) > 0$$

and so it follows that (39) has a non-zero optimum. □

Problem (39) is a non-definite quadratic program, and so is difficult to solve directly. Instead, we convert it into a tractable number of linear programs, using the geometric constructions introduced in §4, as follows.

Lemma 25. *The problem* (39) *has non-zero optimum iff at least one of the following set of problems has a non-zero optimum*

$$P_v : \quad \max_{c \in \widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)} \max_{q \in L_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] \qquad \text{where } v \text{ iterates over the extremal points of } \mathbb{M}(G) \,. \tag{40}$$

Proof. We have $\mathbb{R}^{V \cup E} = \cup_{v \in \operatorname{Ext}(\mathbb{M}(G))} N_{\mathbb{M}(G)}(v)$, as $\mathbb{M}(G)$ is a polytope. Therefore if (c, q, v) is optimal for (39), then it is optimal for P_v, where $v \in \operatorname{Ext}(\mathbb{M}(G))$ is such that $c \in N_{\mathbb{M}}(G)(v)$. Conversely, the optimal value for (39) is the maximum of the optimal values for each problem P_v for $v \in \operatorname{Ext}(\mathbb{M}(G))$. □

Lemma 26. *We can remove the optimization over* $\mathbb{M}(G)$ *in each problem* (40)*; that is, the problem* P_v *is equivalent to*

$$P'_v : \quad \max_{c \in \widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)} \max_{q \in \mathbb{L}_{\frac{\mathrm{u}}{\mathrm{v}}}(G)} \left[\langle c, q \rangle - \langle c, v \rangle \right] \,. \tag{41}$$

Proof. Note that

$$\max_{c \in \widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)} \max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] = \max_{c \in \widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)} \left[\max_{q \in \mathbb{L}_4(G)} \langle c, q \rangle - \max_{x \in \mathbb{M}(G)} \langle c, x \rangle \right] \,.$$

Now observe for $c \in N_{\mathbb{M}(G)}(v)$, we have $\max_{x \in \mathbb{M}(G)} \langle c, x \rangle = \langle c, v \rangle$. □

The problem P'_v in (41) still has a quadratic objective, but has some particularly nice properties that will allow us to make progress. Firstly, it is now purely a maximization problem. Secondly, the region of optimization is the direct product of two polytopes, and the only quadratic terms in the objective are "cross-terms" between variables in these two polytopes.

Lemma 27. *An optimum* $(c^\star, q^\star)$ *of* P'_v *(see* (41)*) occurs with* $c^\star$ *(resp.* $q^\star$*) extremal in* $\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)$ *(resp.* $\mathbb{L}_4$*.)*

Proof. Let $(c^\star, q^\star) \in \widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v) \times \mathbb{L}_4(G)$ be optimal for P'_v. Consider the objective of P'_v with $q^\star$ fixed as a function of the potential parameter c. The resulting objective is linear in c, and so we can move $c^\star$ to some extremal point of $\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)$ without reducing the score attained with $q^\star$ fixed. Therefore we may take $c^\star$ to be extremal. By an analogous argument holding $c^\star$ fixed, we may take $q^\star$ to be extremal. □

The above lemma tells us that to establish whether P'_v has a non-zero optimum, it suffices to check whether any of the following linear programs $P''_{v,c}$ have a non-zero optimum:

$$P''_{v,c} : \quad \max_{q \in \mathbb{L}_4(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] \,, \tag{42}$$

for each $v \in \operatorname{Ext}(\mathbb{M}(G))$, and for each $c \in \operatorname{Ext}(\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v))$. Each problem $P''_{v,c}$ is a linear program over $\mathbb{L}_4(G)$, and is therefore efficiently solvable.

Note that the number of problems $P''_{v,c}$ we have to solve scales with the size of the set $\operatorname{Ext}(\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v))$, and so for computational reasons it is preferable to select the polytope P containing the origin in its interior so that the polytope $\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}(G)}(v)$ has as few vertices as possible.

One way in which to do this is to consider a H-representation for $N_{\mathbb{M}(G)}(v)$ as $\{x \in \mathbb{R}^{V \cup E} | Ax \leq \mathbf{0}\}$ (note that as $N_{\mathbb{M}(G)}(v)$ is a cone based at the origin, all halfspaces intersect the origin, so we may conclude that the linear inequalities defining its H-representation are all of the form $A_i^\top x \leq 0$), and consider the vector $(-\sum_i A_{ij})_j$, enforcing the additional linear constraint that the inner product of all vectors in the convex body with this vector are not too large (say less than or equal to 1). This ensures that we obtain a polytope with number of vertices equal to the number of unique (up to scalar multiplication) extremal rays of the cone (plus an additional vertex at the origin).

11.2 Exploiting Symmetries of the Marginal Polytope and Sherali-Adams Relaxations

Whilst we have reduced the problem (39) to finding the greatest optimal value in a set of linear programs (42), we note that a priori there are a vast number of such LPs that must be solved. Given a graph $G = (V, E)$, there are $2^{|E|}$ possible signings of the graph, whilst there are $2^{|V|}$ vertices of the marginal polytope, and therefore in general many vertices of $\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}}(v)$. Therefore, if possible, we would like to make arguments to reduce the number of linear programs that must be solved to establish whether (39) has a non-zero optimum or not. Fortunately, the rich symmetry groups associated with the marginal polytope and Sherali-Adams relaxations, as discussed in §4, allow large computational savings to be made.

From the remarks in §4, $\mathbb{L}_4(G)$ is not tight for some potential $c \in \mathbb{R}^{V \cup E}$ iff it is not tight for $F(c)$ for any $F \in S_V \rtimes \mathbb{Z}_2^{|V|}$, the symmetry group of the Sherali-Adams polytope. Note that as well as acting naturally on the space of potentials, the symmetry group $S_V \rtimes \mathbb{Z}_2^{|V|}$ of the polytope $\mathbb{L}_4(G)$ also acts naturally on the set of signings of G. From this, we deduce that a given signing Σ of G is non-tight iff $F(\Sigma)$ is non-tight for any (all) $F \in S_V \rtimes \mathbb{Z}_2^{|V|}$. Therefore it suffices to check the optimal value of Problem (42) only for one representative signing of G in each orbit under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$.

Finally, since we have an explicit H-representation of the polytope $\widetilde{\Sigma} \cap P \cap N_{\mathbb{M}}(v)$, we can find a V-representation (i.e. an exhaustive collection of extremal points), and hence recover a list of all LPs of the form of Problem (42) that we need to solve in order to find the optimum of Problem (38). We now detail the results of these calculations for each of the four graphs G of interest i.e. the four graphs in T_4, the minimal forbidden minors for treewidth ≤ 3. The statement of Theorem 12 follows by considering the four minimal forbidden minors for treewidth 3 in turn.

11.3 Forbidden Signed Versions of K_5

We initially calculate the orbits of the set of signings of K_5 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$, as discussed in §11.2. There are 7 different orbits of signings of K_5 - one representative of each is shown in Figure 12:

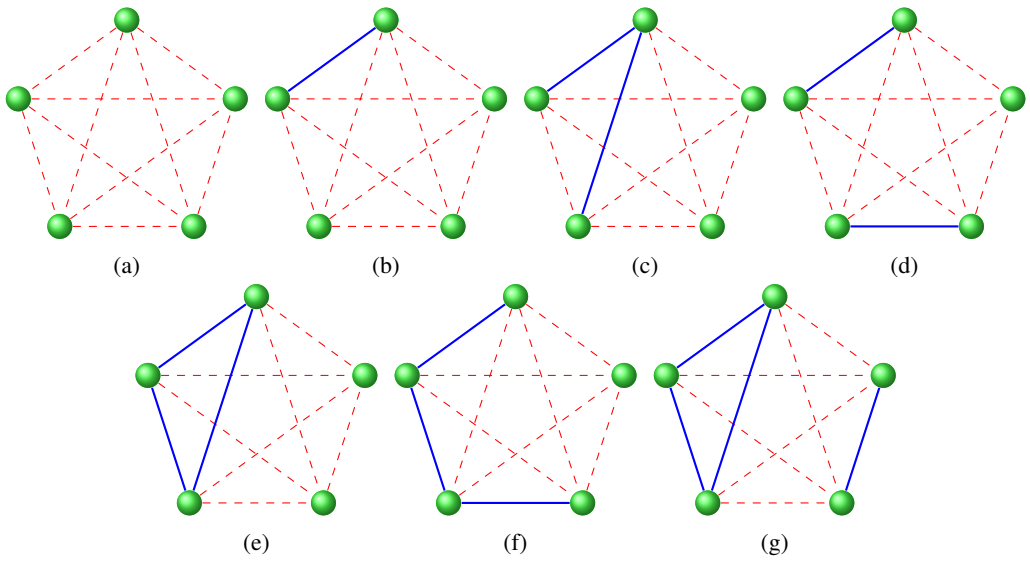


Figure 12: One representative of each of the 7 orbits of signings of K_5 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$. Solid blue edges are attractive and dashed red edges are repulsive.

The tightness of some classes can be determined by known results:

- Figure 12g corresponds to balanced signings; it is known that $\mathbb{L}_2(K_5)$ is tight for such potentials, and so it follows that $\mathbb{L}_4(K_5)$ is too, as $\mathbb{L}_4(G) \subseteq \mathbb{L}_2(G)$.
- Figures 12d and 12e are almost balanced, so are tight for $\mathbb{L}_3(K_5)$ (by Weller et al. (2016)) and so are tight for $\mathbb{L}_4(K_5)$ too, as $\mathbb{L}_4(G) \subseteq \mathbb{L}_3(G)$.

Checking the relevant LPs of the form (42) for the remaining signings reveals that the signings shown in Figures 12b, 12c and 12f are tight with respect to $\mathbb{L}_4(K_5)$, whilst Figure 12a is not. We therefore deduce that the only forbidden signing of K_5 for $\mathbb{L}_4(K_5)$ is that in Figure 12a: the odd-K_5.

11.4 Forbidden Signed Versions of the Octahedral Graph

We initially calculate the orbits of the set of signings of the octahedral graph O_6 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$, as discussed in §11.2. There are 14 different orbits of signings of O_6 - one representative of each is shown in Figure 13:

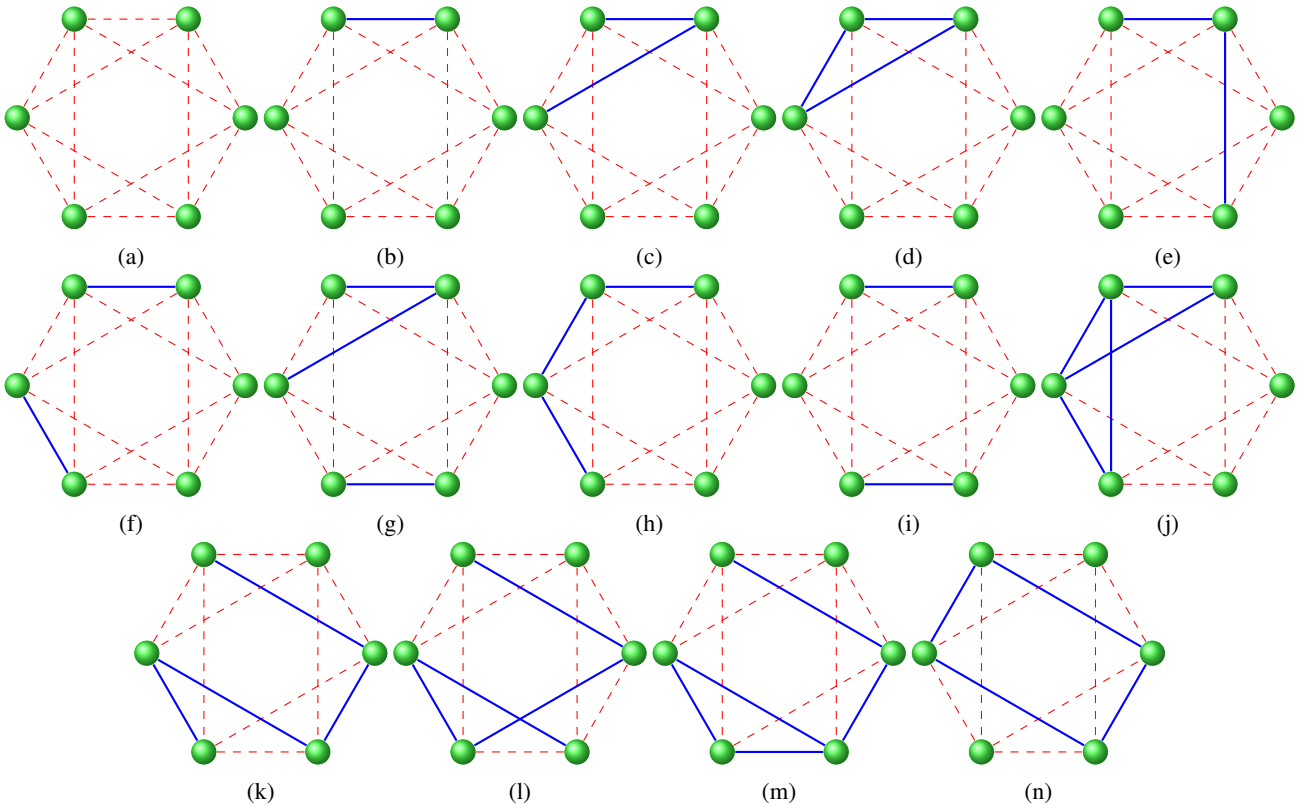


Figure 13: One representative of each of the 14 orbits of signings of O_6 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$. Solid blue edges are attractive and dashed red edges are repulsive.

Checking the relevant LPs of the form (42) for all 14 signings reveals that all signings are tight with respect to $\mathbb{L}_4(O_6)$, and it immediately follows that $\mathbb{L}_4(O_6)$ is tight for any potentials on O_6, and so O_6 need not be forbidden as a minor of a graph G to ensure tightness of $\mathbb{L}_4(G)$.

11.5 Forbidden Signed Versions of the Wagner Graph

We initially calculate the orbits of the set of signings of the Wagner graph M_8 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$, as discussed in §11.2. There are 8 different orbits of signings of M_8 - one representative of each is shown in Figure 14:

Checking the relevant LPs of the form (42) for all 8 signings reveals that all signings are tight with respect to $\mathbb{L}_4(M_8)$, and it immediately follows that $\mathbb{L}_4(M_8)$ is tight for any potentials on M_8, and so M_8 need not be forbidden as a minor of a graph G to ensure tightness of $\mathbb{L}_4(G)$.

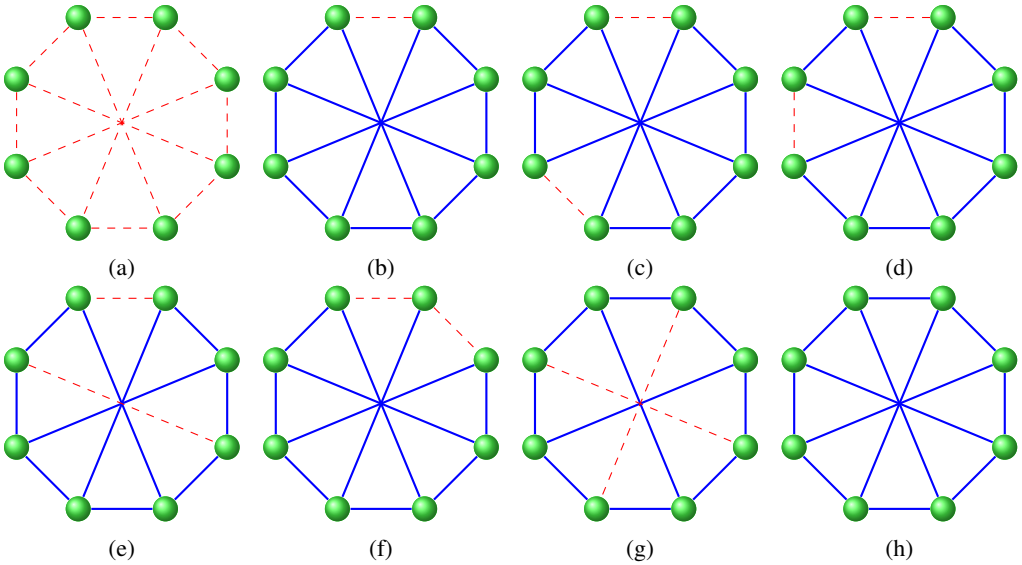


Figure 14: One representative of each of the 8 orbits of signings of M_8 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$. Solid blue edges are attractive and dashed red edges are repulsive.

11.6 Forbidden Signed Versions of the Pentagonal Prism Graph

We initially calculate the orbits of the set of signings of the pentagonal prism graph Y_5 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$, as discussed in §11.2. There are 8 different orbits of signings of Y_5 - one representative of each is shown in Figure 15:

Checking the relevant LPs of the form (42) for all 14 signings reveals that all signings are tight with respect to $\mathbb{L}_4(Y_5)$, and it immediately follows that $\mathbb{L}_4(Y_5)$ is tight for any potentials on Y_5, and so Y_5 need not be forbidden as a minor of a graph G to ensure tightness of $\mathbb{L}_4(G)$.

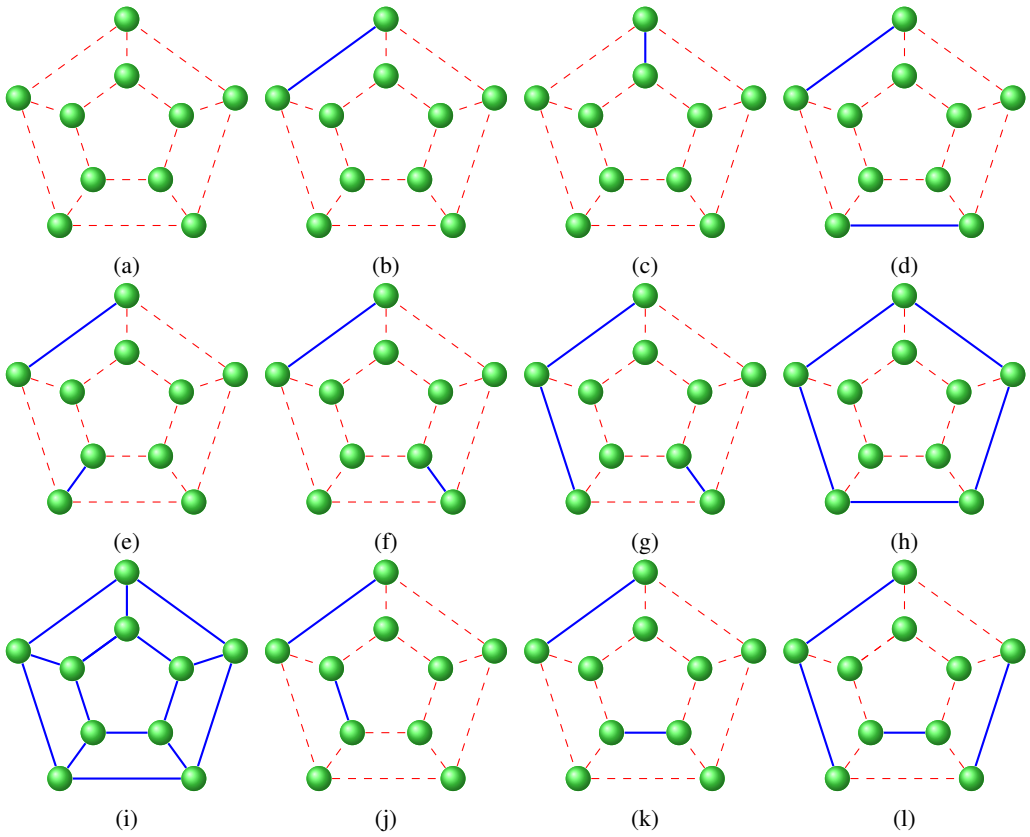


Figure 15: One representative of each of the 12 orbits of signings of Y_5 under the action of $S_V \rtimes \mathbb{Z}_2^{|V|}$. Solid blue edges are attractive and dashed red edges are repulsive.