Abstract

Linear programming (LP) relaxations are widely used to attempt to identify a most likely configuration of a discrete graphical model. In some cases, the LP relaxation attains an optimum vertex at an integral location and thus guarantees an exact solution to the original optimization problem. When this occurs, we say that the LP relaxation is tight. Here we consider binary pairwise models and derive sufficient conditions for guaranteed tightness of (i) the standard LP relaxation on the local polytope LP+LOC, and (ii) the LP relaxation on the triplet-consistent polytope LP+TRI (the next level in the Sherali-Adams hierarchy). We provide simple new proofs of earlier results and derive significant novel results including that LP+TRI is tight for any model where each block is balanced or almost balanced, and a decomposition theorem that may be used to break apart complex models into smaller pieces. An almost balanced (sub-)model is one that contains no frustrated cycles except through one privileged variable.

1 INTRODUCTION

Undirected graphical models, also called Markov random fields (MRFs), are a compact and powerful way to model dependencies among variables, and have become a central tool in machine learning. A fundamental problem is to identify a configuration of all variables that has highest probability, termed maximum a posteriori (MAP) inference. For discrete graphical models, this is a classical combinatorial optimization problem. A popular approach is to express the problem as an integer program, then to relax this to a linear program (LP). If the LP is solved over the convex hull of marginals corresponding to all global settings, termed the marginal polytope, then this would solve the original problem (since an LP attains an optimum at a vertex). However, the marginal polytope has exponentially many facets [Deza and Laurent 2009], hence this space is typically relaxed to the local polytope (LOC), which enforces only pairwise consistency using a linear number of constraints, which we term the LP+LOC approach. If this identifies an optimum at an integer location, then this must be an optimum of the original problem and we say that LP+LOC is tight. [Sontag et al. 2008] demonstrated that using higher-order cluster constraints to tighten LOC to a more constrained yet still tractable polytope, enables many real world examples to be exactly solved efficiently. Using triplets, i.e. clusters of size 3, which leads to the triplet-consistent polytope (TRI), is already very helpful. However, aside from purely topological conditions bounding treewidth, to date there has been little theoretical understanding of when these cluster methods will be effective. In this paper, we focus on binary pairwise models, and provide an important contribution by proving that LP+TRI is guaranteed to be tight for the significant class of models that satisfy the hybrid condition (combining restrictions on topology and potentials) that each block is almost balanced (see §2 for definitions).

We provide background and preliminaries in §2 then in §3 begin by analyzing LP+LOC. By applying a simple yet powerful primal perturbation argument, we first provide new, short proofs of earlier results, then go on to derive novel results on how the optimum varies if one particular marginal is held to various values. These may have applications in other domains, e.g. they may be incorporated into the method of [Weller and Jebara 2014] to yield more efficient approximation of the Bethe partition function. Next in §4 we consider the triplet-consistent polytope TRI. A significant result is that LP+TRI is tight for any model which is almost balanced. In §5 we provide a general decomposition result which may be of independent interest. By combining this with the result in §4 we are able to extend and demonstrate that LP+TRI is tight for any model in which every block is almost balanced. This result is of additional interest since [Weller 2015b] recently demonstrated that a different ‘MWSS’ approach can be used for efficient MAP inference for any valid potentials iff each block of a model is almost balanced. We are able to show that LP+TRI dom-
inates that approach, in the sense that it is guaranteed to be able to solve a superset of models for any potentials.

This subject area has received considerable attention from several communities. We discuss related work throughout the text; for a more comprehensive treatment, see (Wainwright and Jordan [2008] § 8) or (Deza and Laurent [2009]). Note that for binary models (with potentials of any arity), Sontag (2010) has shown that TRI is equivalent to the cycle polytope, which enforces consistency over all cycles.

2 PRELIMINARIES

For binary pairwise models, MAP inference may be framed in a minimal representation (Wainwright and Jordan [2008]) as the discrete optimization problem to identify

\[ x^* \in \text{arg max}_{x \in \{0,1\}^n} \left( \sum_{i \in V} \theta_i x_i + \sum_{(i,j) \in E} W_{ij} x_i x_j \right), \]  

where the model’s topology is given by the graph \((V, E)\), with \(n = |V|\) variables and \(m = |E| \leq \binom{n}{2}\) edge relationships between the variables. The \(n \theta_i\) singleton parameters and \(m W_{ij}\) edge weights define the potentials, and may take any real value. Sometimes we assume all \(\binom{n}{2}\) edges \((i, j)\) are present, allowing for some to have zero weight \(W_{ij} = 0\), where the context will make this clear. Whenever discussing the topology of a model, we mean the graph \((V, E)\).

If \(W_{ij} \geq 0\), the edge \((i, j)\) tends to pull \(X_i\) and \(X_j\) toward the same value and is called attractive. If \(W_{ij} < 0\), the edge is repulsive. We may concatenate the potential parameters together into a vector \(w \in \mathbb{R}^d\), where \(d = n + m\). Similarly, we may define \(y_{ij} = x_i x_j\), then concatenate the \(n x_i\) and \(m y_{ij}\) terms into a vector \(z = (x_1, \ldots, x_n, y_{ij}, \ldots) \in \{0,1\}^d\). This yields the following equivalent integer programming formulation, to identify

\[ z^* \in \text{arg max}_{z \in \{0,1\}^n} w \cdot z \]  

The convex hull of the \(2^n\) possible integer solutions is called the marginal polytope \(\mathbb{M}\). Regarding the convex coefficients as a probability distribution \(p\) over all possible states, \(\mathbb{M}\) may be considered the space of all singleton and pairwise mean marginals that are consistent with some global distribution \(p\) over the \(2^n\) states, that is

\[ \mathbb{M} = \{ z = (z_1, \ldots, z_n, z_{12}, z_{13}, \ldots, z_{(n-1)n}) \} \]  

\[ \text{s.t. } \exists p : z_i = \mathbb{E}_p(X_i) \forall i, \quad z_{ij} = \mathbb{E}_p(X_i X_j) \forall (i, j). \]

A standard approach is to relax (2) to a linear program (LP). However, this remains intractable over \(\mathbb{M}\) (we use tractable to mean solvable in polynomial time) since the number of facets (and hence the number of LP constraints) grows extremely rapidly with \(n\) (Deza and Laurent [2009]). Hence, a

simpler, relaxed constraint set is typically employed, yielding an upper bound on the original optimum. This set is often chosen as the local polytope (LOC or \(L\)), defined as the polytope over \(q = (q_1, \ldots, q_n, \ldots, q_{ij}, \ldots) \in \mathbb{R}^d\) subject to the following linear constraints (see Figure 2):

\[ 0 \leq q_i \leq 1 \quad \forall i \in V, \]  

\[ \max(0, q_i + q_j - 1) \leq q_{ij} \leq \min(q_i, q_j) \quad \forall (i, j) \in E. \]

It is easily checked that these are exactly the requirements to ensure that \(q\) gives rise to valid singleton and pairwise marginals (nonnegative values summing to 1) that are locally consistent (marginalizing a pairwise marginal yields the appropriate singleton marginal), given by

\[ q(X_i = 0) = 1 - q_i, \quad q(X_i = 1) = q_i, \]  

\[ q(X_i = 0, X_j = 0) \quad q(X_i = 0, X_j = 1) \quad q(X_i = 1, X_j = 0) \quad q(X_i = 1, X_j = 1) \]

\[ = \left( 1 + q_{ij} - q_i - q_j - q_{ij} \right). \]

Hence, \(\mathbb{M} \subseteq L\) though \(q \in L\) may not be consistent with any global probability distribution, thus \(q\) is termed a pseudomarginal vector. \(L\) is defined by a polynomial number of constraints, thus it is tractable (Schrijver [1998]) to solve the relaxation LP+LOC given by

\[ q^* \in \text{arg max}_{q \in L} \left( \sum_{i=1}^n \theta_i q_i + \sum_{(i,j) \in E} W_{ij} q_{ij} \right) = \text{arg max } w \cdot q \]  

If an optimum vertex is achieved at an integer solution, then this must be an optimum of the original discrete problem (2), in which case we say that the relaxation is tight.

Starting with LOC, an intuitively appealing series of successively more restrictive relaxations was established by Sherali and Adams [1990]. At order \(r\), the \(L_r\) polytope enforces consistency over all clusters of variables of size \(r\). Hence, \(L_2\) is the local polytope LOC. Next, \(L_3\) enforces consistency over all triplets of variables, which we denote by TRI, and so on. Since \(L_n = \mathbb{M}\), it is clear that \(LP+L_n\) is tight. Building on the junction tree theorem (Cowell et al. 1999), Wainwright and Jordan [2004] demonstrated that a topological sufficient condition for \(LP+L_r\) to be tight, is if a model has treewidth \(\leq r - 1\). Note that this holds for any potentials, whereas looser requirements may suffice given certain restrictions on the potential functions, as we shall show in later Sections.

2.1 Flipping, Balanced and Almost Balanced Models, Block Decomposition and the MWSS Method

If a model has only attractive edges, it is an attractive model, whereas a general model may have any edge types.

\footnotetext{The treewidth of a graph is one less than the smallest possible size of a largest clique in a triangulation of the graph. As examples: a tree has treewidth 1; an \(n \times n\) grid has treewidth \(n\).}
indicating attractive (repulsive) edges. In the balanced block, flipping either \(\{x_1, x_2, x_3\}\) or \(\{x_4, x_5, x_6\}\) partition renders the block attractive. The almost balanced block adds \(x_7\) creating frustrated cycles. On the right, each color indicates a different block of a graph; multi-colored vertices are cut vertices (if these are removed, the graph becomes disconnected), hence belong to multiple blocks.

Figure 1: The left two figures show example model blocks (maximal 2-connected components), with solid blue (dashed red) edges indicating attractive (repulsive) edges. In the balanced block, flipping either \(\{x_1, x_2, x_3\}\) or \(\{x_4, x_5, x_6\}\) partition renders the block attractive. The almost balanced block adds \(x_7\) creating frustrated cycles. On the right, each color indicates a different block of a graph; multi-colored vertices are cut vertices (if these are removed, the graph becomes disconnected), hence belong to multiple blocks.

If a model is not attractive, in some cases it is still possible to render it attractive by flipping (sometimes called switching) a subset of variables, as follows. Partition the variable indices into two subsets, \(A \subseteq [n] = \{1, \ldots, n\}\) and \(B = [n] \setminus A\). Consider the model with new variables \(Y_1, \ldots, Y_n\) where \(Y_i = X_i \forall i \in A\), and \(Y_i = 1 - X_i \forall i \in B\). As described in [Weller 2015a, §2.4], new potential parameters \(\{\theta_i', W_{ij}'\}\) may be determined such that the scores over states are unchanged up to a constant (and hence the distribution is unchanged). In particular, edge weights \(W_{ij}' = \pm W_{ij}\), where the sign changes if exactly one of \(X_i\) and \(X_j\) is flipped. Harary [1953] showed that \(\exists\) a subset \(A \subseteq [n]\) such that flipping those variables renders the model attractive iff there is no cycle with an odd number of repulsive edges. Such a cycle is called a frustrated cycle. Checking for a frustrated cycle may be performed efficiently, and models without frustrated cycles are called balanced. Thus, many results that apply to attractive models may be extended to the wider class of balanced models.

An interesting approach to MAP inference was introduced by [Lebarbier 2009], via a reduction to the maximum weight stable set (MWSS) problem on a derived weighted graph (see Diestel [2010] for all terms from graph theory). Weller [2015b] considered binary pairwise models and proved that this method is guaranteed to yield an efficient optimum configuration for any valid potentials (because the derived graph is perfect) iff each block of the model is almost balanced. A block is a maximal 2-connected subgraph, thus a graph may be repeatedly broken apart at cut vertices to yield its unique block decomposition. A (sub-)model is almost balanced if it may be rendered balanced by deleting one variable (hence, in particular, a balanced (sub-)model is almost balanced). Checking to see if all blocks of a model are almost balanced may be performed efficiently [Weller 2015b]. Our new results show that LP-4-TRI dominates this MWSS approach, see [5]. Figure 1 shows examples of a balanced block, an almost balanced block, and block decomposition.

3 RESULTS FOR LOCAL POLYTOPE

The following perturbation argument will be central in our analysis. Recall that an optimum of an LP is always attained at a vertex (extreme point) of the polytope (Schrijver [1998]). Suppose we wish to show that an optimum vertex may be found with certain properties. Toward contradiction, suppose that all optimum vertices do not have the properties and let \(q^*\) be any such vertex. We shall explicitly construct \(q^+\) and \(q^-\) which lie in the polytope under consideration, such that \(q^* = \frac{1}{2} (q^+ + q^-)\), hence \(q^*\) is not a vertex, and the result follows. We hope this approach might also build understanding by explicitly demonstrating a direction in which the score is nondecreasing. To construct appropriate \(q^+\) and \(q^-\), we shall typically perturb the singleton marginals by symmetric small distances from \(q^*\), and the difficulty will be to ensure that the edge marginal terms can also be perturbed symmetrically.

For the local polytope LOC, given singleton terms \(\{q_i\}\), all pairwise terms \(\{q_{ij}\}\) may be optimized independently. From the constraints (5), optimum edge terms are

\[
q^*_{ij}(q_i, q_j) = \begin{cases} 
\min(q_i, q_j) & \text{if } W_{ij} > 0 \\
\max(0, q_i + q_j - 1) & \text{if } W_{ij} < 0
\end{cases}
\]

Figure 2 indicates the feasible range of \(q_{ij}\) values for ways that \(q_i\) and \(q_j\) might vary together.

Problem cases. If optimum edge terms \(q_{ij}\) are always recomputed, then if \(q_i\) is perturbed up then down by \(\epsilon\), while \(q_j\) is moved by \(\epsilon\) in the same way, in contrary direction or not at all, then the edge term \(q_{ij}\) will always move symmetrically, except in the following two problem cases: (i) \(q_i = q_j\) with an attractive edge \(W_{ij} > 0\), in which case we call \(i\) and \(j\) locked, and \(q_i\) and \(q_j\) must move together; or (ii) \(q_i = 1 - q_j\) with a repulsive edge \(W_{ij} < 0\), in which case we call \(i\) and \(j\) anti-locked, and \(q_i\) and \(q_j\) must move in opposite directions. Observe that case (ii) may be seen as equivalent to case (i) after flipping either variable, see [2] or [8] in the Appendix for more comments on symmetry.
The results in [3,1] were shown previously by other methods [Padberg, 1989], but we provide new, intuitive, short proofs. We believe results in [3,2] and thereafter are now new.

3.1 New Short Proofs of Earlier Results for LOC

Theorem 1. For an attractive model, LP+LOC is tight.

Proof. Toward contradiction, suppose all optimum vertices have some non-integer coordinate. Let \( q^* \in \mathbb{L} \) be such an optimum vertex. Let \( I = \{ i : q^*_i \notin \{0, 1\} \} \). From (6), \( \forall (i,j) \in E, q^*_{ij} = \min(q^+_i,q^-_j) \). Define \( q^+ = (q^+_1, \ldots, q^+_n, \ldots, q^+_n, \ldots) \) as follows:

\[
q^+_i = \begin{cases} 
q^*_i + \epsilon & i \in I \\
q^*_i & i \notin I 
\end{cases}
\]

Note these are optimum edge terms. Similarly, define \( q^- \),

\[
q^-_i = \begin{cases} 
q^*_i - \epsilon & i \in I \\
q^*_i & i \notin I 
\end{cases}
\]

Then we may take any \( \epsilon < \min(a,b) \).

Here \( \epsilon > 0 \) is sufficiently small such that both \( q^+, q^- \in \mathbb{L} \).

More precisely, let \( a = \min_{i \in I} q^*_i, b = \min_{i \in I} (1-q^*_i) \) then we may take any \( \epsilon < \min(a,b) \).

It is easily checked that \( q^* = \frac{1}{2} (q^+ + q^-) \), hence \( q^* \) is not a vertex.

The particular choice of \( q^+ \) and \( q^- \) in the proof above works by ensuring that all edge terms \( \{q_{ij}\} \) move symmetrically, i.e. each edge term either does not move for both \( q^+ \) and \( q^- \), or moves up for one and down for the other.

Theorem 2. For a balanced model, LP+LOC is tight.

Proof A. Since the model is balanced, a subset of variables may be identified such that flipping them renders the model attractive [Harary, 1953], see (2.1); then apply Theorem 1, if a model is tight then so too is any flipping of it.

\[ F^\lambda(x) = \max_{q \in F^\lambda_q = x} w \cdot q, \quad x \in [0, 1]. \]

First we provide the following simple Lemma.

Lemma 5. For any \( \mathbb{P}, F^\lambda(x) \) is a concave function for \( x \in [0, 1] \).

Proof. Given any \( x_0, x_1 \in [0, 1] \), let \( q^0, q^1 \in \mathbb{R}^d \) be arg max locations for \( F^\lambda(x_0) \) and \( F^\lambda(x_1) \) respectively. For any \( \lambda \in [0, 1] \), let \( \bar{x} = \lambda x_1 + (1-\lambda)x_0 \) and \( \bar{q} = \lambda q^1 + (1-\lambda)q^0 \). Now \( F^\lambda(\bar{x}) = \max_{q \in F^\lambda_q = \bar{q}} w \cdot q \geq w \cdot \bar{q} = \lambda F^\lambda(x_1) + (1-\lambda)F^\lambda(x_0) \).

More generally, going forward, we may typically take any positive \( \epsilon < \min \) distance to any of the values that we claim all singleton marginals must take (here these values are 0 and 1).
Using Theorem 4 and Lemma 5 we shall show how $F^*_L(x)$, the constrained optimum on LOC, varies with $x$.

**Theorem 6.** For a balanced model, $F^*_L(x)$ is linear.

**Proof.** We assume an attractive model. The result will then extend to a balanced model by first flipping an appropriate subset of variables, see Equation (2.1). We shall show here that $F^*_L(x)$ is convex, then linearity follows from Lemma 3.

For any $y \in [0, 1]$, consider an arg max of $F^*_L(y)$ as given by Theorem 4. Partition the variables into 3 exhaustive sets: $A_y = \{ j : q_j = 0 \}, B_y = \{ j : q_j = y \}$ and $C_y = \{ j : q_j = 1 \}$. Define the function $f_y : [0, 1] \rightarrow \mathbb{R}$ given by $f_y(x) = f(q(x); y)$ where $q(x; y)$ is defined by:

$$q_j(x; y) = \begin{cases} 0 & j \in A_y \\ x & j \in B_y \\ 1 & j \in C_y \end{cases}$$

using optimum terms $q_{jk}(x; y) = \min \{ q_j(x; y), q_k(x; y) \}$ for all edges. Observe that $f_y(x)$ is the linear function achieved by holding fixed the partition of variables $A_y, B_y, C_y$ that was determined for the arg max of the constrained optimum at $q_i = y$. Now $F^*_L(x) = \sup_{y \in [0, 1]} f_y(x)$, hence is convex. \qed

Note that since $F^*_L(x)$ is linear, it must be that each of the linear $f_y(x)$ functions from the proof are equal, so as an immediate corollary, we may take the $A, B, C$ sets to be constant with the same variables in them, independent of $y$.

For a general model, we can show an analog of Theorem 4.

**Theorem 7.** For a general model, if one variable’s marginal $q_i = x \in [0, 1]$ is fixed and we optimize over all others $\{ q_j : j \neq i \}$, then an optimum is achieved with $q_j \in \{ 0, x, \frac{1}{2}, 1 - x, 1 \}$ \forall $j$.

**Proof.** Fix $q_i = x$ and optimize over all other variables. Let $\mathcal{I} = \{ j : q_j \notin \{ 0, x, \frac{1}{2}, 1 - x, 1 \} \}$. If $\exists j \in \mathcal{I}$, take $A$ to be all variables in $\mathcal{I}$ equal to $q_j$ and $B$ to be all variables in $\mathcal{I}$ equal to $1 - q_j$. Perturb $A$ and down $B$, then vice versa, i.e. set $q^+$ and $q^-$ as in (7). \qed

Observe that (because of the fixed $\frac{1}{2}$ in its statement) Theorem 7 does not allow an argument as in the proof of Theorem 6 to yield the (false) conclusion that $F^*_L(x)$ is linear for a general model.

### 4 RESULTS FOR TRIPLET POLYTOPE

The triplet-consistent polytope TRI is defined by the constraints of the local polytope $L$ (3), together with the following additional triangle inequalities (4 per triplet):

$$\forall i < j < k, \quad q_i + q_{jk} \geq q_{ij} + q_{ik}, \quad (9)$$

$$q_{ij} + q_{ik} + q_{jk} \geq q_i + q_j + q_k - 1. \quad (10)$$

These enforce consistency over any triplet of variables, as may be derived by the lift-and-project method. Hence, $\mathbb{M} \subseteq \text{TRI} \subseteq \mathbb{L}$. For the purpose of these inequalities, if an edge $(i, j) \notin \mathcal{E}$ then assume it is present with $W_{ij} = 0$. See Appendix 6 for a derivation of the inequalities, and 7 for a discussion of their symmetry.

In this Section, we shall show that, somewhat remarkably, an almost balanced model on TRI behaves in many ways just like a balanced model on LOC. A key result is the following analog of Theorem 1.

**Theorem 8.** For an almost balanced model, LP+TRI (the LP relaxation over TRI) is tight.

To prove Theorem 8, we shall show the following analog of Theorem 6, where $s$ is the special variable in an almost balanced model such that when removed, the remainder is balanced (see (2.1)).

**Theorem 9.** For an almost balanced model with special variable $s$, $F^*_\text{TRI}(x)$ is a linear function.

If we can prove Theorem 9 then an optimum occurs at $s = 0$ or $s = 1$. Conditioning on this value of $s$ yields a balanced model; then Theorem 8 follows by Theorem 2 (since TRI $\subseteq \mathbb{L}$).

For our perturbation method, on LOC, once we condition on a set of singleton marginals, the edge marginals are independent and easily computed. On TRI, in contrast, edges interact. We call any edge where the optimum edge marginal takes its maximum possible value on LOC (behaving ‘like an attractive edge’, though the edge may be repulsive), a strong up edge. Similarly, we call an edge where the optimum marginal takes its minimum possible value on LOC (behaving ‘like a repulsive edge’), a strong down edge. Generalizing from (3) 2 variables are locked up (locked down) if they have $q_i = q_j$ ($q_i = 1 - q_j$) and are joined by a strong up (strong down) edge; in either case (up or down) the edge is locking. A cycle of strong (up or down) edges is strong frustrated if it contains an odd number of strong down edges.

**Problem triangles.** In addition to the earlier problem cases for LOC in (3) involving 2 variables, from which we observe that if we have locked up (locked down) variables, they must move together (opposite), we identify the following four new ‘problem triangles’ (see Appendix 9 for details) over 3 variables in TRI for our perturbation method, i.e. cases where perturbing singleton marginals up and down by a small $\epsilon$ will not lead to symmetric changes in edge marginals. Each form has 3 strong edges and is strong frustrated: (i) One strong down edge $b - c$ with $b + c < 1$ and $a = b + c$, see Figure 3; (ii) One strong down edge $b - c$ with $b + c > 1$ and $a = b + c - 1$; (iii) Three strong down edges $a + b + c = 1$ (this implies that each pair sum to
less than 1); (iv) Three strong down edges with

\[
q_{ab} + q_{ac} = q_a
\]

lower bound
for \( q_{bc} \)

Figure 3: Above: an illustration of ‘problem triangle’ type (i). Blue edges are strong up, the red wavy edge is strong down. Below: a plot showing the relevant triangle constraint (others are always satisfied) \( q_a + q_{bc} \geq q_{ab} + q_{ac} \) as \( q_a \) is varied, holding fixed \( q_b \) and \( q_c \) while recomputing LOC-optimum edge marginals for \( q_{ab} \) and \( q_{ac} \). The TRI constraint is binding where the plot is red, and not where it is black. Here we consider \( q_a + q_c < 1 \), hence on LOC, \( q_{bc} = 0 \), and \( q_{ab} = \min(q_a, q_b) \), \( q_{ac} = \min(q_a, q_c) \). \( q_a = q_b + q_c \) is the new problem case (e.g. if just \( q_a \) is perturbed, the constraint becomes binding just on one side leading to an asymmetric response). There may also be problems at \( q_a \in \{ \min(q_a, q_b), \max(q_a, q_c) \} \) but these are already covered since they would form locking edges from \( a \) to \( b \) or \( c \).

Lower case letters such as \( a \) may be overloaded for variable names and their singleton marginals.

\[
q_{ab} + q_{ac} = q_a
\]

not binding

\[
\text{constraint binding}
\]

\[
\begin{array}{c}
\text{lower bound} \\
\text{for } q_{bc}
\end{array}
\]

\[
\begin{array}{c}
0 \\
\min(q_a, q_c) \\
\max(q_a, q_c) \\
q_a
\end{array}
\]

To simplify analysis, without loss of generality, by flipping an appropriate set of variables in \( V \setminus \{ s \} \) (see [2,1]), we may assume that we have an ‘almost attractive’ model, with all edges attractive, except for some edges incident to \( s \); results then extend to almost balanced models.

With these observations, we first provide a key Lemma on the structure of strong down edges.

**Lemma 10.** In an almost attractive model with special variable \( s \) (i.e. the model on \( V \setminus \{ s \} \) is attractive), if all edge marginals have been optimized in TRI given a set of singleton marginals, then any strong down edge must connect via a path of strong down edges to \( s \) in one of two particular ways: either all edges on the path have sum of incident variable marginals \( < 1 \) with form shown in Figure 4 or all edges have sum of incident variable marginals \( > 1 \) with form shown in Figure 7 (in Appendix). Considering only strong down edges, there can be no odd cycles (which rules out problem triangles (iii) and (iv)). Further, any variable on a problem triangle of form (i) or (ii), which is opposite the strong down edge, cannot be incident to any strong down edge in the model.

**Proof.** This is a consequence of applying the TRI constraints to various triangles, details in Appendix 11.

Using Lemma 10, we show an analog of Theorem 4 which will enable the subsequent proof of Theorem 9.

**Theorem 11.** In an almost balanced model with special
TRI is \( \frac{1}{2} \)-integral for \( n \leq 5 \), but as \( n \) grows, vertices of TRI at fractions with arbitrarily large denominator are possible.

5 MODEL DECOMPOSITION RESULTS

In this section we show a general result that an LP relaxation of a component-structured graphical model is tight whenever the LP relaxations on the components are tight and consistency is enforced on the variables in common between adjacent components. Consider a graphical model with variables \( V = A \cup B \), and let \( C = A \cap B \) be the variables in common between \( A \) and \( B \). Specifically, let \( p(\vec{x}, \vec{y}, \vec{z}) \) be an exponential family distribution with sufficient statistic vector \( \phi(\vec{x}, \vec{y}, \vec{z}) = \left[ \phi_x(\vec{x}), \phi_y(\vec{y}), \phi_z(\vec{z}) \right] \), and let \( A = X \cup Y \) and \( B = Y \cup Z \).

Let \( M \) be the marginal polytope corresponding to \( \phi(\vec{x}, \vec{y}, \vec{z}) \), i.e. the convex hull of \( \phi(\vec{x}, \vec{y}, \vec{z}) \) for every assignment to \( X, Y, Z \). Similarly, let \( M_A \) and \( M_B \) be the marginal polytopes corresponding to sufficient statistic vectors \( \left[ \phi_x(\vec{x}), \phi_y(\vec{y}), \phi_z(\vec{z}) \right] \), respectively. Every polytope can be equivalently defined as the intersection of linear inequalities (the polytope’s maximal facets). Let \( M_I = M_A \cap M_B \) be the polytope defined by combining the linear inequalities making up both \( M_A \) and \( M_B \).

**Theorem 12** (Decomposition result for graphical models). Suppose we have two polytopes \( M_A \) and \( M_B \) for models with variables \( A \) and \( B \), where \( C = A \cap B \) are the variables in common. Suppose we have LP relaxations for \( M_A \) and \( M_B \) which are known to be tight for any objective \( \theta_A \in \Theta_A \) and \( \theta_B \in \Theta_B \), respectively. If the sets \( \Theta_A \) and \( \Theta_B \) are closed under the addition of an arbitrary potential function \( \theta_C \), then \( M_I = M_A \cap M_B \) (defined just above) is tight on the combined model over variables \( A \cup B \), i.e. \( M_I = M \).

**Proof.** Clearly \( M_I \) is a polytope and \( M \subseteq M_I \), i.e. \( M_I \) is a relaxation, which we shall demonstrate is tight. We do this by showing that for every weight vector \( \vec{w} \), the optimal value of \( \vec{w} \cdot \mu \) is the same for \( \mu \in M_I \) as for \( \mu \in M \). To do that, we consider the Lagrangian relaxation and demonstrate a dual witness.

For any \( \vec{w} = \left[ w_x, w_y, w_z \right] \), let \( \theta_{\vec{w}}(\vec{x}, \vec{y}, \vec{z}) = \vec{w} \cdot \phi(\vec{x}, \vec{y}, \vec{z}) = \theta_y(\vec{x}, \vec{y}) + \theta_z(\vec{x}, \vec{z}) \), where \( \theta_y(\vec{x}, \vec{y}) = \left[ w_x, w_y \right] \cdot \left[ \phi_x(\vec{x}), \phi_y(\vec{y}) \right] \) and \( \theta_z(\vec{x}, \vec{z}) = \left[ 0, w_z \right] \cdot \left[ \phi_x(\vec{x}), \phi_z(\vec{z}) \right] \). Consider the following:

\[
\max_{\vec{x},\vec{y},\vec{z}} \theta(\vec{x}, \vec{y}, \vec{z}) = \max_{\vec{w} \in M} \vec{w} \cdot \mu \leq \max_{\mu \in M_I} \vec{w} \cdot \mu
\]

\[
= \max_{\mu_1 \in M_A, \mu_2 \in M_B, \mu(\vec{x}) = \mu_1(\vec{x})} \left[ w_x, w_y \right] \cdot \mu_1 + \left[ 0, w_z \right] \cdot \mu_2
\]

\[
= \min_{\lambda_x} \left( \max_{\mu_1 \in M_A, \mu_2 \in M_B} \left[ w_x, w_y \right] \cdot \mu_1 + \left[ 0, w_z \right] \cdot \mu_2 \right.
\]

\[
+ \left. \lambda_x (\mu_1(\vec{x}) - \mu_2(\vec{x})) \right)
\]
\[
\min_{\lambda x} \left( \max_{\vec{x}, \vec{y}} [\theta_y(\vec{x}, \vec{y}) + \lambda x] + \max_{\vec{x}, \vec{z}} [\theta_z(\vec{x}, \vec{z}) - \lambda z] \right),
\]

where in the last step we use the assumption that \( M_A \) and \( M_B \) are tight for any potential \( \theta(\vec{x}) \).

Now plug in \( \lambda x = \frac{[\max_{\vec{x}, \vec{y}} \theta_y(\vec{x}, \vec{z}) - \max_{\vec{y}} \theta_y(\vec{x}, \vec{y})]}{2} \), and one can verify that the last term is equal to \( \max_{\vec{x}, \vec{y}, \vec{z}} \theta(\vec{x}, \vec{y}, \vec{z}) \), and thus the inequality must be an equality, which proves that the relaxation is tight. \( \square \)

As a special case, for Sherali-Adams relaxations we have

**Corollary 13.** If \( LP+L_r \) (clusters of up to \( r \) variables) is tight for model \( A \), and similarly \( LP+L_s \) is tight for model \( B \), in each case no matter what the single-node potentials are, and with the two models having exactly one variable in common, then \( LP+L_t \) is tight on the combined MRF over all the variables, where \( t = \max(r, s) \).

### 5.1 Application to LP+TRI, Comparison to MWSS Approach

Wainwright and Jordan (2004) showed that LP+TRI is tight for any model that has treewidth \( \leq 2 \). Theorem 8 shows that LP+TRI is tight for any model that is almost balanced. Applying Corollary 13 we deduce that LP+TRI is tight for any model with block structure such that each block is either almost balanced or has treewidth 2 (a model with treewidth 1 is a tree hence is balanced).

An interesting approach to MAP inference was introduced by Jebara (2009) and Sanghavi et al. (2009), which reduces the problem to the graph theoretic challenge of identifying a maximum weight stable set (MWSS) in a derived weighted graph termed a \textit{naive Markov random field} (NMRF). For binary pairwise models, Weller (2015b) demonstrated that this method will yield an exact solution (via a perfect graph) in polynomial time for any valid potentials iff each block of the model is almost balanced.

Our result demonstrates that the LP+TRI approach can handle all these models and more. For example, Figure 6 shows a 2-connected model that is not almost balanced (since it contains two disjoint frustrated cycles \( x_3-x_7-x_8 \) and \( x_5-x_7-x_6 \)), thus it is not always solvable by the MWSS approach. Solid blue (dashed red) edges are attractive (repulsive).

We have provided short, intuitive proofs and derived new results that deepen our understanding and may help to provide guidance in practice, including a general decomposition result (Theorem 12). Theorem 8 on hybrid conditions (combining restrictions on topology and potentials) for tightness of LP+TRI is interesting for several reasons. It improves our understanding of why and when the relaxation will perform well. It supports the interesting characterization of almost balanced models, which, to our knowledge, was not much considered prior to the recent analysis of Weller (2015b). It shows that LP+TRI dominates the MWSS approach, in the sense that LP+TRI is guaranteed to solve a strict superset of MAP inference problems for any potentials in polynomial time. Finally, it provides an important step into hybrid characterizations, which remains an exciting uncharted field following success in characterizations of tractability using only topological constraints (Chandrasekaran et al., 2008), or only families of potentials (Kolmogorov et al., 2015; Thapper and Živný, 2015).

In future work, we plan to examine higher order relaxations in the Sherali-Adams hierarchy, which impose consistency over larger clusters. LP+LOC=\( L_3 \) is tight for any balanced model and we now know that LP+TRI=\( L_3 \) is tight for any almost balanced model. It will be interesting to explore whether LP+\( L_4 \) is tight for any model that can be rendered balanced by deleting two variables.

It may be tempting to conjecture that if LP+\( L_r \) is tight over a model class for some \( r \), then if an extra variable is added with arbitrary interactions, LP+\( L_{r+1} \) will be tight on the larger model. However, this is false. Consider a planar binary pairwise model with no singleton potentials. LP+TRI is tight for such models (Barahona, 1983); yet if one adds a new variable connected to all of the original ones, the MAP inference task becomes NP-hard (Barahona, 1982).

---

**Tightness of LP Relaxations for Almost Balanced Models**

---

**Figure 6:** Illustration of a 2-connected model with treewidth 2, hence LP+TRI is tight for any potentials; but it is not almost balanced (since it contains two disjoint frustrated cycles \( x_3-x_7-x_8 \) and \( x_5-x_7-x_6 \)), thus it is not always solvable by the MWSS approach. Solid blue (dashed red) edges are attractive (repulsive).
References


7 Derivation of the Triangle Inequalities

Here we show how to derive the inequalities defining TRI, i.e. (9) and (10) together with the standard constraints for LOC (3), following the lift-and-project method as described in [Wainwright and Jordan 2008] Example 8.7). We first ‘lift’ to the space of marginals over three variables, where we require that a well-defined probability distribution exists over every triplet of variables in the model. Next we ‘project’ the resulting constraints back down to our familiar space of singleton and pairwise marginals, defined (in the minimal representation) by a vector of dimension \( d = n + m \), where \( n \) is the number of variables, each with a \( q_i \) term, and \( m \) is the number of edges, each with a \( q_{ij} \) term.

Recall that each set of terms \( \{q_i, q_j, q_{ij}\} \), provided they are feasible in LOC, defines a valid probability distribution on the pair of variables \( q_i, q_j \) as shown in (4), which we reproduce here:

\[
\begin{align*}
q(X_i = 0, X_j = 0) & \quad q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & \quad q(X_i = 1, X_j = 1)
\end{align*}
\]

Observe that 4 terms are required for a distribution over variables \( X_i \) and \( X_j \), but given \( \{q_i, q_j\} \), we have several constraints: all must sum to 1, which leaves 3 degrees of freedom; then in order to match the singleton marginals given by \( q_i \) and \( q_j \), this removes 2 more degrees of freedom leaving just one, which here is specified by \( q_{ij} \). Note that enforcing that all terms are nonnegative yields the LOC inequalities (3).

Similarly, when considering a distribution over 3 variables, say \( i, j \) and \( k \), there are 8 terms but given \( \{q_i, q_j, q_k, q_{ij}, q_{ik}, q_{jk}, q_{ijk}, \alpha\} \), we must satisfy the following constraints: all must sum to 1, marginalizing out any one variable must give the appropriate pairwise term (3 constraints), and marginalizing out any two variables must give the appropriate singleton term (3 constraints). Thus just one free term remains (in fact, it is not hard to see that for a cluster of any size, there is always just one degree of freedom, given all lower order terms), which here we shall specify using \( \alpha = q_{ijk} = q(X_i = 1, X_j = 1, X_k = 1) \).

Given \( \{q_i, q_j, q_k, q_{ij}, q_{ik}, q_{jk}, \alpha = q_{ijk}\} \), it is straightforward to see that we may write down the probabilities of all 8 states as follows:

With \( k = 0 \),

\[
\begin{align*}
q(X_i = 0, X_j = 0) & \quad q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & \quad q(X_i = 1, X_j = 1)
\end{align*}
\]

\[
\left(1 - q_i - q_j - q_k + q_{ij} + q_{ik} + q_{jk} - \alpha\right) \quad q_j + \alpha - q_{ij} - q_{ik} \quad q_{ij} - \alpha
\]

With \( k = 1 \),

\[
\begin{align*}
q(X_i = 0, X_j = 0) & \quad q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & \quad q(X_i = 1, X_j = 1)
\end{align*}
\]

\[
\left(q_k + \alpha - q_{ik} - q_{jk} \quad q_{jk} - \alpha\right)
\]

We have the inequalities that all 8 expressions must be nonnegative. Now to project back down to our original space, \( \alpha \) must be eliminated, which can be achieved using Fourier-Motzkin elimination [Schrijver 1998] as follows: (i) first express all inequalities in \( \leq \) form with \( \alpha \) (the variable to be eliminated) isolated; then (ii) combine \( \leq \alpha \) constraints with \( \alpha \leq \) constraints in pairs in order to yield a new inequality without \( \alpha \).
Working through this algebra yields exactly the constraints of LOC and TRI, i.e. (3), (9) and (10). As one example, to obtain the first inequality of (9), which is that \( q_i + \alpha - q_{ij} - q_{ik} \geq 0 \) \( \iff q_{ij} + q_{ik} - q_i \leq \alpha \), with the inequality from the top right of the lower matrix, i.e. \( q_{jk} - \alpha \geq 0 \) \( \iff \alpha \leq q_{jk} \).

8 Symmetry: Flipping, Polytope Constraints and Problem Triangles

The minimal representation can sometimes obscure the underlying symmetry of the system. We demonstrate that the constraints for each of the local and triplet polytopes may be obtained by starting with just one constraint then flipping variables and applying the constraint to the flipped models. (This illustrates the symmetry but note that it is not true that having all constraints is the same as having just one constraint.)

Suppose we have a model including variables \( X_i \) and \( X_j \) with an edge \((i, j)\) between them, together with a pseudo-marginal vector \( q \). If \( X_i \) is flipped then we consider the model with \( Y_i = 1 - X_i \) and \( Y_j = X_j \). Let the new equivalent pseudo-marginal vector be \( q' \). Clearly \( q'_i = 1 - q_i \) and \( q'_j = q_j \). For the edge marginal, observe that

\[
\begin{align*}
\text{Original edge marginal} & \quad (q(X_i = 0, X_j = 0) \quad q(X_i = 1, X_j = 0)) = (1 + q_{ij} - q_i - q_j \quad q_i - q_{ij}), \\
\text{New edge marginal} & \quad (q'(Y_i = 0, Y_j = 0) \quad q'(Y_i = 1, Y_j = 0)) = (1 + q'_i - q'_j - q'_i \quad q'_i - q'_{ij}).
\end{align*}
\]

To equate terms, note that \( Y_i = 1 \) or \( 0 \) corresponds to \( X_i = 0 \) or 1, so the row order has been reversed. Hence, \( q'_{ij} = q_j - q_{ij} \).

The constraints that \( 0 \leq q_i \leq 1 \ \forall i \in \mathcal{V} \), and \( 0 \leq q_{ij} \leq 1 \ \forall (i, j) \in \mathcal{E} \) are base constraints that hold without considering multiple variables.

8.1 Local Polytope LOC

Let us start with the following one constraint (other choices would also work),

\[ q_{ij} \leq q_i. \]

Flipping \( X_i \) and applying the above constraint to the new model yields

\[ q'_i \leq q'_i \quad \iff \quad q_j - q_{ij} \leq 1 - q_i \quad \iff \quad q_{ij} \geq q_i + q_j - 1. \]

Now take the last constraint above and flip \( X_j \) to obtain

\[ q_i - q_{ij} \geq q_i + 1 - q_j - 1 \quad \iff \quad q_{ij} \leq q_j. \]

Observe that we have obtained all the local polytope constraints.

8.2 Triplet Polytope TRI

Consider any triplet of variables \( X_i, X_j, X_k \). Let us start with the following one constraint,

\[ q_i + q_{jk} \geq q_{ij} + q_{ik}. \]

Flip \( X_i \) to obtain

\[ 1 - q_i + q_{jk} \geq q_j - q_{ij} + q_k - q_{ik} \quad \iff \quad q_{ij} + q_{jk} + q_{ik} \geq q_i + q_j + q_k - 1. \quad (11) \]

Take the last constraint above and flip \( X_j \) to obtain

\[ q_i - q_{ij} + q_k - q_{jk} + q_{ik} \geq q_i + 1 - q_j + q_k - 1 \quad \iff \quad q_j + q_{ik} \geq q_{ij} + q_{jk}. \]

Instead, take (11) and flip \( X_k \) to obtain

\[ q_{ij} + q_j - q_{jk} + q_i - q_{ik} \geq q_i + q_j + 1 - q_k - 1 \quad \iff \quad q_k + q_{ij} \geq q_{ik} + q_{jk}. \]

Observe that all the triplet polytope constraints may be obtained.
Figure 7: A triangle abc with two attractive edges a−b and a−c, and one repulsive edge b−c, together with a graph of the relevant triangle constraint \(q_a + q_{bc} \geq q_{ab} + q_{ac}\) as \(q_a\) is varied, holding fixed \(q_b\) and \(q_c\) while recomputing LOC-optimum edge marginals for \(q_{ab}\) and \(q_{ac}\). The constraint is binding where the plot is red, and not where it is black. Here we consider \(q_b + q_c < 1\), hence on LOC, \(q_{bc} = 0\), and \(q_{ab} = \min(q_a, q_b), q_{ac} = \min(q_a, q_c)\). Observe that \(q_a = q_b + q_c\) is the one new case that causes trouble (e.g. if just \(q_a\) is perturbed, the constraint becomes binding just on one side leading to an asymmetric response). With these one can also be difficulties at the vertices at \(q_a\) or \(q_{bc}\), hence are already covered by the LOC cases. When \(q_a = q_b + q_c\), observe that any sufficiently small perturbation of singleton marginals up and down by a vector within the following two dimensional space will work symmetrically for edge marginals: \((\delta a, \delta b, \delta c) = \alpha(1,1,0) + \beta(1,0,1)\). In particular, this includes a small perturbation of \((\delta a, \delta b, \delta c) = \pm \epsilon(0,1,-1)\).

8.3 Symmetry of Problem Triangles in TRI

Consider Figure 7. If \(q_b + q_c < 1\) and \(q_a = q_b + q_c\), with \(a−b\) and \(a−c\) strong up edges and \(b−c\) a strong down edge, then this is a problem triangle of type (i) as described in \[9\] it has 3 strong edges with \(a−b\) and \(a−c\) strong up, and \(b−c\) strong down; in addition, \(b + c < 1\) and \(a = b + c\). We shall show that the other 3 types of problem triangle described in \[4\] may be obtained from this one by flipping variables.

The following observations are easily checked:
Flipping a variable flips each of its incident edges between strong up ↔ strong down.

Since flipping variables always changes an even number of edges, any flipping of our original problem triangle yields a triangle with three strong edges including an odd number of strong down edges, i.e. a strong frustrated triangle.

First, flip \(a\) to yield a triangle with 3 strong down edges and singleton marginals \(a’ = 1−a, b’ = b, c’ = c\). Now \(a = b + c \Leftrightarrow a’ + b’ + c’ = 1\), i.e. problem triangle type (iii). Note that we have \(b’ + c’ = a < 1\); also \(a = b + c\) hence \(a > b\) and \(a > c\), which implies that \(a’ + b’ < 1\) and \(a’ + c’ < 1\).

Now flip all variables to give \(a’’ = 1−a’, b’’ = 1−b’, c’’ = 1−c’\). This again yields a triangle with 3 strong down edges but now \(a’’ + b’’ = 1−a’+1−b’ > 1\), and similarly \(a’’ + c’’ > 1, b’’+c’’ > 1\). We have \(a’’ + b’’ + c’’ = 1−a’+1−b’+1−c’ = 2\), i.e. problem triangle type (iv).

Finally, flip \(a’’\) to yield \(a’’’ = 1−a’, b’’’ = 1−b’, c’’’ = 1−c’\), forming a strong triangle with edges incident to \(a’’’\) strong up and \(b’’’−c’’’\) strong down. Now \(a’’’ + b’’’ + c’’’ = 2 \Leftrightarrow 1−a’’’+b’’’+c’’’ = 2 \Leftrightarrow a’’’ = b’’’ + c’’’ − 1, with \(b’’’ + c’’’ > 1\), i.e. problem triangle type (ii).

9 Problem Cases for TRI

It is easily checked that the TRI constraints (9)-(10) are always satisfied by any triangle abc unless there is a sufficient degree of ‘frustration’. Each of the 3 constraints of the form (9) prevent one of the edges from being too low relative to the other two. These prevent, for example, a triangle where \(q_a = q_b = q_c = \frac{1}{2}\) with two strong up edges \(q_{ab} = q_{ac} = \frac{1}{2}\) and one strong down edge \(q_{bc} = 0\), though this is feasible in LOC. The constraint (10) prevents, for example, a triangle with the same singleton marginals having three strong down edges \(q_{ab} = q_{bc} = q_{ac} = 0\), though this is also feasible in LOC.

We shall consider a triangle on \(a, b, c\) with \(q_b + q_c < 1\), and the TRI constraint \(q_a + q_{bc} \geq q_{ab} + q_{ac}\). Other cases follow by symmetry, see \[8\]. We may assume that all \(a, b, c \in (0, 1)\) (see \[10, 2\]). Consider the illustrations in Figure 7 and the text in the caption below it. TRI-optimum edges may be considered as LOC-optimum edges, adjusted as required to satisfy...
TRI constraints. For our perturbation method, we are concerned with when symmetric changes in singleton marginals result in asymmetric changes in edge marginals. This occurs when there is some asymmetric kink in a constraint. As previously observed, the LOC constraints cause trouble for locking edges. Regarding Figure 4, we see that there are 3 potentially problematic locations. As discussed in the caption text, the new one is when \( q_i = q_j = q_k \), corresponding to a type (i) problem triangle (see Figure 4 for descriptions of the different types of problem triangle). To maintain this relationship and avoid trouble, any perturbation in the caption text will work, i.e. \( (\delta a, \delta b, \delta c) = \alpha(1,1,0) + \beta(1,0,1) \) for sufficiently small \( \alpha, \beta > 0 \). In particular, this includes a small perturbation of \( (\delta a, \delta b, \delta c) = \pm \varepsilon(0,1,-1) \). By similar reasoning, or symmetry, a similar result is obtained for type (ii) problem triangles. For type (iii) and (iv) problem triangles, where all edges are strong down, instead a sufficiently small perturbation of any 2 of the variables in opposite directions works. Note that in all cases, if exactly 2 of the variables that are adjacent by a strong down edge are perturbed in sufficiently small opposite directions, while leaving the third variable untouched, this works.

10 Locking Components, and 0 or 1 Singleton Marginals

We first analyze locking components, see §10.2 for variables with 0 or 1 singleton marginals.

10.1 Locking Components

On TRI, given marginals \( q_i, q_j, q_{ij} \), we say that variables \( i \) and \( j \) are locked up if \( q_i = q_j \) and \( q_{ij} = \min(q_i, q_j) \), i.e. they have the same singleton marginal and there is a strong up edge between them. Similarly, we say that variables \( i \) and \( j \) are locked down if \( q_i = 1 - q_j \) and \( q_{ij} = \max(0, q_i + q_j - 1) \), i.e. they have ‘opposite’ singleton marginals and there is a strong down edge between them. In either case, we say that the edge \((i, j)\) is locking (either up or down).

We say that a cycle is strongly frustrated if it is composed of strong edges with an odd number of strong down edges.

Define a locking component to be a component of the model that is connected when considering only locking edges. This means that between any 2 variables in the locking component, there exists some path between them composed only of locking edges. In general, this path might be long but the next result shows that in TRI, in fact it is always of length 1. In addition, we see that a locking component contains no strong frustrated cycle.

**Lemma 14.** In TRI, within any locking component, all pairs of variables are adjacent via locking edges; further, there are no strong frustrated triangles, and hence no strong frustrated cycles.

**Proof.** For the first part, the following result is sufficient, since given a path between any 2 variables in the component, this will allow us iteratively to find a path shorter by one edge, until we get the edge directly between them:

Suppose variable \( A \) is adjacent to \( B \) which is adjacent to \( C \), each via a locking edge. We shall show that \( A \) is adjacent to \( C \) via a locking edge so as always to avoid a strong frustrated triangle. Let \( B \) have singleton marginal \( x \). We shall consider all marginals, where \( A \) means singleton marginal for \( A \) etc., \( AB \) means edge marginal for edge \( A - B \) etc. There are 3 cases:

1. \( A - B \) is locking up, \( B - C \) is locking up. \( A : x, B : x, C : x, AB : x, BC : x \). Now triangle inequality \( B + AC \geq AB + BC \) gives \( AC = x \), i.e. \( A - C \) is locking up.
2. \( A - B \) is locking up, \( B - C \) is locking down. \( A : x, B : x, C : 1 - x, AB : x, BC : 0 \). Now \( A + BC \geq AB + AC \) gives \( AC = 0 \), i.e. \( A - C \) is locking down.
3. \( A - B \) is locking down, \( B - C \) is locking down. \( A : 1 - x, B : x, C : 1 - x, AB : 0, BC : 0 \). Now \( AB + BC + AC \geq A + B + C - 1 \) gives \( AC = 1 - x \), i.e. \( A - C \) is locking up.

We have shown that all variables in the locking component are adjacent via locking edges, and that no triangle is strongly frustrated. To demonstrate that there are no strong frustrated cycles (of any length): Suppose toward contradiction that there exists such a cycle, and let us pick one with minimum length composed of variables \( v_1, v_2, \ldots, v_n \), so \( n \geq 4 \) is minimal. Now ‘break’ the cycle into two pieces: \( \{v_1, v_2, \ldots, v_{n-1}\} \) and \( \{v_{n-1}, v_n, v_1\} \). Since the second piece is a triangle, by the above it is not strongly frustrated, i.e. the number of strong down edges in it is \( 0 \mod 2 \). Edge \( v_1 - v_{n-1} \) is either strongly up or strongly down, either way, twice the number of its strong down edges is \( 0 \mod 2 \). Let \( r \) be the number of strong down edges in cycle \( v_1, v_2, \ldots, v_{n-1} \mod 2 \), then we have \( r + 0 = 1 \mod 2 \), contradiction since \( n \) was minimal. \( \square \)
we have

\[ \frac{\partial f}{\partial \hat{q}_{ij}} = \frac{\partial f}{\partial q_{ij}} - \frac{\partial f}{\partial q_{ik}} - \frac{\partial f}{\partial q_{jk}} \]

\[ \frac{\partial f}{\partial \hat{q}_{jk}} = \frac{\partial f}{\partial q_{jk}} - \frac{\partial f}{\partial q_{ik}} - \frac{\partial f}{\partial q_{ij}} \]

\[ \frac{\partial f}{\partial \hat{q}_{ik}} = \frac{\partial f}{\partial q_{ik}} - \frac{\partial f}{\partial q_{ij}} - \frac{\partial f}{\partial q_{jk}} \]

**Lemma 16.** If a variable has singleton marginal 0 or 1, then its incident edge marginals are forced and will move symmetrically (on LOC or TRI). For any triplet containing the variable, all TRI inequalities are always satisfied for any (LOC valid) opposite edge marginal.

**Proof.** This follows directly from the relevant definitions (see \[\sqrt{4}\] and Lemma \[\sqrt{14}\] since a problem triangle has no locking edges.

\[ \begin{aligned}
q_i + q_j &\ge q_{ij} + q_{ik}, \text{ i.e. } x + r \ge x + y, \text{ hence } r \ge y. \\
q_j + q_k &\ge q_{jk} + q_{ij}, \text{ i.e. } x + y \ge x + r, \text{ hence } r \le y.
\end{aligned} \]

**10.1.2 A problem triangle cannot have more than one variable in a specific locking component**

This follows directly from the relevant definitions (see \[\sqrt{4}\] and Lemma \[\sqrt{14}\]) since a problem triangle has no locking edges.

\[ \text{10.2 0 or 1 Singleton Marginals} \]

We consider any variable \( X_i \) with singleton marginal \( q_i \in \{0, 1\} \).

**Lemma 16.** If a variable has singleton marginal 0 or 1, then its incident edge marginals are forced and will move symmetrically (on LOC or TRI). For any triplet containing the variable, all TRI inequalities are always satisfied for any (LOC valid) opposite edge marginal.

**Proof.** If variable \( X_i \) has singleton marginal \( q_i = 0 \), then for any incident edge \((i, j)\), by the LOC constraint \( q_{ij} \le q_i \), we have \( q_{ij} = 0 \). If instead \( X_i \) has singleton marginal \( q_i = 1 \), then for any incident edge \((i, j)\), by the LOC constraint \( q_{ij} \ge q_i + q_j - 1 \), we have \( q_{ij} = q_j \).

Consider any triplet formed by \( X_i \) together with any variables \( X_j \) and \( X_k \), which have singleton marginals \( q_j \) and \( q_k \). Let \( q_{jk} \) be the LOC-valid edge marginal for the edge \( X_j - X_k \) (i.e. \( q_{jk} = q_{ij} = q_{ik} \)). It is straightforward to check that all TRI constraints (given by \[\sqrt{9} - \sqrt{10}\]) are satisfied. We demonstrate this for the case \( q_i = 0 \):

\[ q_i + q_{jk} - q_{ij} - q_{ik} = 0 + q_{jk} - 0 - 0 \]
\[ q_j + q_{ik} - q_{ij} - q_{jk} = q_j + 0 - 0 - q_{jk} \]
\[ q_k + q_{ij} - q_{ik} - q_{jk} = q_k + 0 - 0 - q_{jk} \]
\[ q_{ij} + q_{jk} + q_{ik} - q_i - q_j - q_k + 1 = 0 + q_{jk} + 0 - 0 - q_j - q_k + 1 \]

\[ = q_{jk} \ge 0 \]
\[ = q_j - q_{jk} \ge 0 \]
\[ = q_k - q_{jk} \ge 0 \]

\[ = q_{jk} - (q_j + q_k - 1) \ge 0 \]

\[ \Box \]

**11 The Structure of Strong Down Edges in TRI in an Almost Attractive Model**

Here we consider an almost attractive model with special variable \( s \), i.e. the model on \( \mathcal{V}' = \mathcal{V} \setminus \{s\} \) is attractive, and provide a proof of Lemma \[\sqrt{10}\] (with extensions). We assume that singleton marginals are given and, conditioned on those,
optimum edge marginals have been computed.

On LOC, any edges not incident to \( s \) will take their maximum possible edge marginal, hence if an optimum occurs with any of these being strong down, it must be due to a triangle constraint holding it down (aside from edges incident to variables with singleton marginal of 0 or 1 which are always both strong up and strong down, see [10] we shall ignore such variables for the remainder of this Section).

For simplicity, we shall sometimes overload lower case letters such as \( a \) to represent both variable names and their singleton marginal values (i.e. we sometimes write \( a \) for \( q_a \)), where the context should make it clear which is being used.

**Lemma 17.** On TRI, consider variables \( a, b, c \) with given singleton marginals \( b \in (0, 1) \). If edge \( a - b \) is attractive, then in order for it to be strong down at an optimum, one of the following (mutually exclusive) conditions must hold:

- If \( a + b < 1 \), then
  - Either \( a + c < 1 \), \( a - c \) is a strong down edge, \( b \leq c \) and \( b - c \) is a strong up edge; or
  - (the same situation on the other side) \( b + c < 1 \), \( b - c \) is a strong down edge, \( a \leq c \) and \( a - c \) is a strong up edge.

- If \( a + b > 1 \), then
  - Either \( a + c > 1 \), \( a - c \) is a strong down edge, \( b \geq c \) and \( b - c \) is a strong up edge; or
  - (the same situation on the other side) \( b + c > 1 \), \( b - c \) is a strong down edge, \( a \geq c \) and \( a - c \) is a strong up edge.

**Proof.** We first consider \( a + b < 1 \), and hence for a strong down edge, \( q_{ab} = 0 \). Of the four TRI constraints operating on \( a, b, c \), there are two that upper bound \( q_{ab} \):

\[
(1) \quad q_a + q_{bc} \geq q_{ab} + q_{ac} \quad \text{and} \quad (2) \quad q_b + q_{ac} \geq q_{ab} + q_{bc}.
\]

One of these two inequalities must enforce \( q_{ab} = 0 \), let us assume (2), hence \( q_b + q_{ac} - q_{bc} = 0 \). Since (by LOC constraints) \( q_b - q_{bc} \geq 0 \) and \( q_{ac} \geq 0 \), we must have \( q_{ac} = 0 \) and \( q_{bc} = q_b \). This means we must have \( a + c < 1 \), \( a - c \) is a strong down edge, \( b \leq c \) and \( b - c \) is a strong up edge.

If instead (1) enforces \( q_{ab} = 0 \), then we conclude symmetrically that \( b + c < 1 \), \( b - c \) is a strong down edge, \( a \leq c \) and \( a - c \) is a strong up edge.

If \( a + b > 1 \), arguing as above and assuming (2) enforces the constraint, we have \( q_b + q_{ac} - q_{bc} = q_a + q_b - 1 \). Rewrite this as \((q_{ac} - q_a - q_c + 1) + (q_c - q_{bc}) = 0 \). Similarly to the earlier case, by LOC constraints both terms in parentheses are constrained to be \( \geq 0 \) and the result follows.

In an almost attractive model, only edges incident to \( s \) are repulsive. If these have weak edge potentials, that at an optimum, these edges might be pulled up from their LOC optima. If this happens to all the repulsive edges, then by the above reasoning, there will be no strong down edges in the model. The only way strong down edges can appear at an optimum is by starting with a strong down edge incident to \( s \) and then applying the result in Lemma[17]

In order to get a path of strong down edges from \( s \), we can repeatedly apply Lemma[17] to determine which singleton and edge marginals are possible. Here, let us suppose there is a path \( s - a - b - c - d \) of strong down edges of length 4 (this will indicate the form, which continues for a path of any length). Applying Lemma[17] to triangle \( sab \) then \( abc \) then \( bcd \), we see that for every strong down edge on the path, either the sum of incident variable singleton marginals is always \( < 1 \), or else the sum is always \( > 1 \).

We consider the case \( s + a < 1 \) (and hence along the whole strong down path, the sum of singleton marginals across each strong down edge is \( < 1 \)), see Figure[9] Again by Lemma[17] we see that for each variable along the path starting with \( b \) each variable has a strong up edge to the variable that comes two earlier than it, and that two earlier variable must have a singleton marginal at least as great.

Now consider triangle \( sac \) and apply the TRI inequality \( q_c + q_{sa} \geq q_{cs} + q_{ca} \) to obtain \( c + 0 \geq q_{cs} + c \) and hence, \( c - s \) is a strong down edge with \( q_{cs} = 0 \) (note this is feasible since \( c + s \leq a + s \leq 1 \)). Similar reasoning applied to \( abd \) gives \( d - a \) is a strong down edge with \( q_{ad} = 0 \).

Consider triangle \( sbd \) and apply the TRI inequality \( q_b + q_{ds} \geq q_{bs} + q_{bd} \) to obtain \( b + q_{ds} \geq b + d \), hence \( q_{ds} = d \) (since \( q_{ds} \leq d \) by a LOC constraint).

Similar reasoning shows that for each variable along the strong down path from \( s \), as we move away from \( s \), the edge marginal to \( s \) alternates between 0 (for an odd distance by strong down edges) and the respective singleton marginal (for
Tightness of LP Relaxations for Almost Balanced Models

Figure 9: Illustration of how marginals must behave for an almost attractive model in TRI to obtain a path of strong down edges $s - a - b - c - d$, shown in wavy red, for the case $s + a < 1$ (hence all strong down edges have edge marginal 0). Singleton marginals are shown in black. The lighter pink wavy edges $s - c$ and $a - d$ are implied also to have 0 edge marginal. The other edges (straight blue) are forced to be strong up edges, and show their edge marginal in blue. Note that as we move along the strong down path from $s$, the edge marginal to $s$ alternates between 0 (for an odd distance by wavy edges) and the respective singleton marginal (for an even distance by wavy edges); in particular it is not possible to have an odd cycle composed only of strong down edges. Two problem triangles of type (i) are shown: $x, a, b$ and $y, s, a$. In each case, the variable ($x$ or $y$) opposite the strong down edge necessarily connects to $s$ via an edge marginal that is strictly greater than 0 and less than the variable’s singleton marginal; it is impossible for such a variable also to be incident to a strong down edge.

A problem triangle of type (ii) cannot occur with a strong down edge along this path (since here incident singleton marginals sum to less than 1), but a problem triangle of type (i) can. This is composed of a strong down edge with incident singleton marginals that sum to less than 1, together with an ‘opposite’ variable that connects by strong up edges to either side of the strong down edge, and with singleton marginal equal to the sum of the two other variables’ marginals. See the example problem triangle $xab$ shown in Figure 9. Consider triangle $sbx$: applying $q_b + q_{bx} \geq q_{bs} + q_{bx}$ gives $b + q_{bx} \geq b + b$, hence $q_{bx} \geq b$. Now consider triangle $sax$: applying $q_x + q_{sx} \geq q_{sx} + q_{xa}$ gives $a + b + 0 \geq a + q_{sx}$ (using $x = a + b$), hence $q_{sx} = b$.

Similar reasoning shows that in any problem triangle of type (i), the ‘opposite’ variable must have an edge to $s$ with edge marginal strictly > 0 and less than its own singleton marginal.

If instead we have $s + a > 1$, similar reasoning leads to the analogous result illustrated in Figure 10 where problem triangles of type (ii) are shown.

12 Proof of Theorem 11 from Section 4

Theorem 11. In an almost balanced model with special variable $s$, if we fix $q_s = x \in [0, 1]$ and optimize in TRI over all other marginals, then an optimum is achieved with: $q_j \in \{0, x, 1 - x, 1\} \forall j$; all edges (other than to variables which have 0 or 1 singleton marginal) are locking or anti-locking, with no strong frustrated cycles.

Proof. We shall show that any variables $\notin \{0, 1\}$ that are not locked or anti-locked to $s$ may be perturbed with symmetric edge marginals, demonstrating that we are not at an optimum vertex. As shown earlier, we may assume an almost attractive model with no locking components and no variables $\in \{0, 1\}$. It remains to show that we may successfully perturb all problem triangles. Using the structural result of Lemma 10 and recalling the discussion of problem triangles, we may symmetrically perturb as follows.
Divide variables that are on a strong down edge path from $s$ into two sets: $A$ contains those where the strong down edges have marginal 0 (i.e. the type $s + a < 1$, shown in Figure 9); $B$ contains those where the strong down edge $> 0$ (i.e. the type $s + a > 1$, shown in Figure 10). These are disjoint (consider edge marginals to $s$).

Further divide these into $A_1$, $A_2$, $B_1$, and $B_2$ by odd and even distance from $s$ by strong down edges (in fact it is not hard to see that shortest distances of 1 or 2 are the only possibilities since a longer path implies a shorter path with the same parity).

Let $Y_A$ be all $y$ type variables that lie on a problem triangle opposite a strong down edge $s - a$ with marginal 0. As examples, regarding Figure 9 we have $a \in A_1$ and $y \in Y_A$.

Let $Y_B$ be all $y$ type variables that lie on a problem triangle opposite a strong down edge $s - a$ with marginal $> 0$. As examples, regarding Figure 10 we have $a \in B_1$ and $y \in Y_B$.

Note that $Y_A$ may intersect with $B_2$ (both have an edge marginal of value $s$ to the variable $s$), similarly $Y_B$ may intersect with $A_2$. Given the observations on symmetric perturbations in problem triangles from §9, since we will not perturb $s$, $Y_A$ must move with $A_1$, and because of the possible non-empty intersection $Y_A \cap B_2$, we must move these together with $B_2$; similarly for $Y_B$.

Hence, a symmetric perturbation by sufficiently small $\epsilon > 0$ is achieved as follows, keeping $s$ fixed: for $q^+$, perturb $A_1 \cup Y_A \cup B_2$ up, $B_1 \cup Y_B \cup A_2$ down; for $q^-$ do the opposite.