Tightness of LP Relaxations for Almost Balanced Models

Adrian Weller
University of Cambridge

Mark Rowland
University of Cambridge

David Sontag
New York University

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 existing 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-
Tightness of LP Relaxations for Almost Balanced Models

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), Montag [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

$$x^* \in \arg \max_{x \in \{0,1\}^n} \left( \sum_{i \in V} \theta_i x_i + \sum_{(i,j) \in E} W_{ij} x_i x_j \right),$$

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 may 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 $z = (x_1, x_2, \ldots, x_n, y_{ij}, \ldots) \in \{0,1\}^d$. This yields the following equivalent integer programming formulation, to identify

$$z^* \in \arg \max_{z \in \{0,1\}^n} w \cdot z$$

The convex hull of the $2^n$ possible integer solutions is called the marginal polytope $M$. Regarding the convex coefficients as a probability distribution $p$ over all possible states, $M$ 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

$$M = \{z = (z_1, z_2, z_3, \ldots, z_{(n-1)n}) \text{ s.t. } \exists p : z_i = E_p(X_i) \forall i, z_{ij} = E_p(X_i X_j) \forall (i, j)\}.$$

A standard approach is to relax (2) to a linear program (LP). However, this remains intractable over $M$ (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

$$\text{singletons } q(X_i = 0) = 1 - q_i, \quad q(X_i = 1) = q_i,$$

$$\text{edges } (\begin{pmatrix} q(X = 1, X_i = 0) & q(X_i = 1, X_j = 0) \\ q(X = 0, X_j = 0) & q(X_i = 0, X_j = 1) \end{pmatrix} = \begin{pmatrix} 1 + q_{ij} - q_i - q_j & q_{ij} - q_i q_j \\ q_i - q_{ij} & q_{ij} \end{pmatrix}.$$  

Hence, $M \subseteq L$ though $q \in L$ may not be consistent with any global probability distribution, thus $q$ is termed a pseudo-marginal vector. $L$ is defined by a polynomial number of constraints, thus it is tractable Schrijver [1998] to solve the relaxation $L+LP$ given by

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

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 $L_n = M$, 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.

\footnote{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 $n$.}
If a model is not attractive, in some cases it is still possible to render it attractive by flipping some edges (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 $\{\theta_i', W_{ij}^\prime\}$ 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 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 Jebara (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+TRI dominates this MWSS approach, see §5.1. Figure 1 shows examples of a balanced block, an almost balanced block, and block decomposition.

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.

### 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. This constructively demonstrates a direction in which the score is nondecreasing (a similar approach was used by Taskar et al., 2004). 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

$$q^*_{ij}(q_i, q_j) = \begin{cases} \min(q_i, q_j) & \text{if } W_{ij} > 0 \\ \max(0, q_i + q_j - 1) & \text{if } W_{ij} < 0 \end{cases}. \tag{6}$$

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 §2.1 or §8 in the Appendix for more comments on symmetry.
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 \( \mathcal{I} = \{ i : q^*_i \notin \{0, 1\} \} \). From (6), \( \forall (i,j) \in \mathcal{E}, q_{ij} = \min(q^*_i, q^*_j) \). Define \( q^+ = (q^*_1, \ldots, q^*_n) \) as follows:

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

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

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

Here \( \epsilon > 0 \) is sufficiently small such that both \( q^+, q^- \in \mathbb{L} \). More precisely, let \( a = \min_{i \in \mathcal{I}} q^*_i, b = \min_{i \notin \mathcal{I}} (1 - q^*_i) \) 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).

**Proof B.** We provide an alternative derivation which essentially incorporates the flipping into the proof. Recall the two possible problem cases described above in (3).

Split \( \mathcal{I} = \{ i : q^*_i \notin \{0, 1\} \} \) into two groups, \( A \) and \( B \), such that all intra-group edges are attractive and all inter-group edges are repulsive (flipping either group renders the model attractive, see (2.1)). Observe that \( q^* = \frac{1}{2} (q^+ + q^-) \) if we define

\[
q^+_i = \begin{cases} 
q^*_i + \epsilon & i \in A \\
q^*_i - \epsilon & i \in B, q^-_i \end{cases} \quad q^-_i = \begin{cases} 
q^*_i - \epsilon & i \in A \\
q^*_i + \epsilon & i \in B, i \notin \mathcal{I}
\end{cases}
\]

with both using optimum edge terms \( \{q^+_{ij}, q^-_{ij}\} \).

**Theorem 3.** For a general model (any potentials, attractive or not), LP+LOC is half-integral.

**Proof.** Let \( A = \{ i : 0 < q^*_i < \frac{1}{2} \} \), let \( B = \{ i : \frac{1}{2} < q^*_i < 1 \} \). Set \( q^+ \) and \( q^- \) as in (7), with \( \mathcal{I} = A \cup B \).

Since Theorem 5 considers optimizing an arbitrary linear function over the polytope LOC, an immediate corollary is that all vertices of LOC are half-integral.

3.2 New Results for LOC, Fixing One Variable and Optimizing Over the Others

Results in this Section may be of independent interest, and also serve as a warm-up for our approach for TRI in (4).

**Theorem 4.** For an attractive model, if we fix one variable's marginal \( q_j = x \in [0, 1] \), and optimize over all others \( \{ q_j : j \neq i \} \), then an optimum vertex is achieved with \( q_j \in \{0, x, 1\} \).

**Proof.** Toward contradiction, if all optima have some \( q^*_j \notin \{0, x, 1\} \) then construct \( q^+ \) and \( q^- \) by moving these variables up/down together by \( \epsilon \), i.e. the same construction for \( q^+ \) and \( q^- \) as in the proof of Theorem 1, setting positive \( \epsilon < \min \) distance to any member of \( \{0, x, 1\} \).

We define the following constrained optimum function for any polytope \( \mathcal{P} \) which is a relaxation of \( \mathbb{M} \),

\[
F^i_{\mathcal{P}}(x) = \max_{q_j \in \mathcal{P}, q_i = x} w \cdot q, \quad x \in [0, 1].
\]

First we provide the following simple Lemma.

**Lemma 5.** For any \( \mathcal{P} \), \( F^i_{\mathcal{P}}(x) \) 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 maxes for \( F^i_{\mathcal{P}}(x_0) \) and \( F^i_{\mathcal{P}}(x_1) \) respectively. For any \( \lambda \in [0, 1], \) let \( \tilde{x} = \lambda x_1 + (1 - \lambda) x_0 \) and \( \tilde{q} = \lambda q^1 + (1 - \lambda) q^0 \). Now \( F^i_{\mathcal{P}}(\tilde{x}) = \max_{q_j \in \mathcal{P}, q_i = x} w \cdot q \geq w \cdot \tilde{q} = \lambda F^i_{\mathcal{P}}(x_1) + (1 - \lambda) F^i_{\mathcal{P}}(x_0) \).
Using Theorem 4 and Lemma 5 we shall show how \( F^L_i(x) \), the constrained optimum on LOC, varies with \( x \).

**Theorem 6.** For a balanced model, \( F^L_i(x) \) 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 [2.1]. We shall show here that \( F^L_i(x) \) is convex, then linearity follows from Lemma 5.

For any \( y \in [0, 1] \), consider an \( \arg \max \) of \( F^L_i(y) \) 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 \( f_y : [0, 1] \to \mathbb{R} \) given by 

\[
 f_y(x) = f(q(x); y) \quad \text{where} \quad q(x; y) \text{ is defined by:}
\]

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

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 \( F^L_i(x) = \sup_{y \in [0, 1]} f_y(x) \), hence is convex.

Note that since \( F^L_i(x) \) is linear, it must be that each of the linear \( f_y(x) \) functions from the proof are equal, so 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 \( q_i \) is \( x \in [0, 1] \) 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 \( I = \{ j : q_j \notin \{0, x, \frac{1}{2}, 1 - x, 1\} \} \). If \( \exists j \in I \), take \( A \) to be all variables in \( I \) equal to \( q_j \) and \( B \) to be all variables in \( I \) equal to \( 1 - q_j \). Perturb \( A \) up and \( B \) down, then vice versa, i.e. set \( q^+ \) and \( q^- \) as in (7).

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 \( F^L_i(x) \) 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):

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

These enforce consistency over any triplet of variables, as may be derived by the lift-and-project method. Hence, \( M \subseteq TRI \subseteq L \). For the purpose of these inequalities, if an edge \( (i, j) \notin E \) then assume it is present with \( W_{ij} = 0 \). See Appendix [7] for a derivation of the inequalities, and [8] 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 \( s \), \( F^s_{TRI}(x) \) 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 \( \subseteq L \)).

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 [12, 1] 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 with \( a + b + c = 1 \) (this implies that each pair sums to

\[
 q_{ij} + q_{ik} + q_{jk} \geq q_i + q_j + q_k - 1. \quad (10)
\]
Lower case letters such as a may be overloaded for variable names and their singleton marginals.

\begin{align*}
q_{ab} + q_{ac} &= q_a \\
\text{lower bound for } q_{bc}
\end{align*}

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 \( q_a \in \{ \min(q_a, q_c), \max(q_a, q_c) \} \) but these are already covered since they would form locking edges from \( a \) to \( b \) or \( c \).

Less than 1); (iv) Three strong down edges with \( a+b+c = 2 \) (this implies that each pair sums to more than 1). In each of the four cases, only certain combined perturbations of variables will result in symmetric edge marginal perturbations. In all cases, it works if exactly 2 (of the 3) variables are perturbed, to move in opposite directions, with the 2 variables being on either end of a strong down edge. We shall see that this is always sufficient for an almost balanced model.

Locked (up or down) variables must move appropriately. See Appendix [2] for details of the following: Variables connected by paths of locking edges form in TRI a locking component, in which all variables are adjacent by locking edges and there is no strong frustrated cycle. If we know the edge marginal from any member of a locking component to a variable outside it, we can uniquely determine all edge marginals (which move together/opposite) to that outside variable from all members of the locking component. A problem triangle cannot have more than one variable in a specific locking component. Hence, we may ‘contract’ any locking component to a single variable for analysis purposes and assume we have no locking components. If any variable has singleton marginal 0 or 1, then this uniquely determines incident edge marginals, which will always satisfy the TRI constraints and move symmetrically. Hence we may assume no variables with 0 or 1 singleton marginal.

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 [10].

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
A minimal example of a block that is not almost balanced, also a minimal example of a block that has treewidth \( \geq 2 \), hence models with this topology might not be tight for LP+TRI. Solid blue (dashed red) edges are attractive (repulsive). All triangles are frustrated with an odd number of repulsive edges.

\[ x_1 \]
\[ x_2 \]
\[ x_3 \]

![Figure 5: A minimal example of a block that is not almost balanced, also a minimal example of a block that has treewidth \( \geq 2 \), hence models with this topology might not be tight for LP+TRI. Solid blue (dashed red) edges are attractive (repulsive). All triangles are frustrated with an odd number of repulsive edges.](image)

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 above, we may assume an almost attractive model with no locking components and no variables \( \in \{0, 1\} \). Using the structural result of Lemma 10 we may construct a symmetric perturbation, as required, see Appendix 11 for details. \( \Box \)

Using Theorem 11, Theorem 12 may be proved in the same way as was shown for Theorem 6 (recall Lemma 5 applies this by showing that for every weight vector \( \tilde{θ} \) and \( \tilde{ϕ} \), there exists \( \tilde{ϕ}_q(\tilde{x}, \tilde{y}, \tilde{z}) \) such that \( \tilde{ϕ}_q(\tilde{x}, \tilde{y}, \tilde{z}) = \tilde{ϕ}_q(\tilde{x}, \tilde{y}, \tilde{z}) \cdot \tilde{ϕ}_q(\tilde{x}, \tilde{y}, \tilde{z}) \).

### 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 \( p(\bar{x}, \bar{y}, \bar{z}) \) be an exponential family distribution with sufficient statistic vector \( \phi(\bar{x}, \bar{y}, \bar{z}) = [\phi_a(\bar{x}), \phi_y(\bar{y}), \phi_z(\bar{z})] \), and let \( A = X \cup Y \) and \( B = Y \cup Z \).

Let \( M \) be the marginal polytope corresponding to \( \phi(\bar{x}, \bar{y}, \bar{z}) \), i.e. the convex hull of \( \phi(\bar{x}, \bar{y}, \bar{z}) \) for every assignment to \( X, Y, Z \). Similarly, let \( M_A \) and \( M_B \) be the marginal polytopes corresponding to sufficient statistic vectors \( \phi_a(\bar{x}) \) and \( \phi_y(\bar{y}) \), 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 \( \tilde{θ} \), the optimal value of \( \tilde{w} \cdot \tilde{θ} \) is the same for \( \tilde{w} \in M_I \) as for \( \tilde{w} \in M \). To do that, we consider the Lagrangian relaxation and demonstrate a dual witness.

For any \( \tilde{w} \) \( [w_{xy}, w_{yx}, w_{zx}] \), let \( \theta_A(\bar{x}, \bar{y}, \bar{z}) = w_{xy} \cdot \phi(\bar{x}, \bar{y}, \bar{z}) = \theta_y(\bar{x}, \bar{y}) + \theta_z(\bar{x}, \bar{z}) \), where \( \theta_y(\bar{x}, \bar{y}) = [w_{xy}, w_{yx}] \cdot [\phi_y(\bar{y}), \phi_y(\bar{y})] \) and \( \theta_z(\bar{x}, \bar{z}) = [0, w_{zx}] \cdot [\phi_y(\bar{y}), \phi_y(\bar{y})] \). Consider the following:

\[
\max_{x, \bar{y}, \bar{z}} \theta(\bar{x}, \bar{y}, \bar{z}) = \max_{x, \bar{y}, \bar{z}} \tilde{w} \cdot \tilde{θ} \leq \max_{x, \bar{y}, \bar{z}} \tilde{w} \cdot \tilde{θ} = \max_{\tilde{w} \in M_I} \tilde{w} \cdot \tilde{θ} = \max_{\tilde{w} \in M} \tilde{w} \cdot \tilde{θ} = \max_{\tilde{w} \in M} \tilde{w} \cdot \tilde{θ}.
\]

### 4.1 Remarks

A minimal example of a block that is not almost balanced is shown in Figure 5. If there are no singleton potentials, then by the analysis of the cut polytope by Barahona (1983), TRI = \( \mathbb{R}_+^n \) and hence LP+TRI is tight. However, potentials exist s.t. LP+TRI is not tight for models with this topology.

**Non-integral Vertices of TRI.** Padberg (1989) proved that LOC = \( \mathbb{L}_3 \) is \( \frac{1}{2} \)-integral (we showed a new, short, proof, see Theorem 5), and also showed that TRI = \( \mathbb{L}_3 \) has no \( \frac{1}{3} \)-integral vertex. Hence, the triangle inequalities are sufficient to cut off all fractional vertices of LOC. It is natural to wonder if perhaps TRI = \( \mathbb{L}_3 \) is \( \frac{1}{3} \)-integral. Laurent and Poljak (1995) considered this by analyzing the metric polytope (which is equivalent via the covariance mapping, Hammel 1965, Deza 1973, De Simone 1989/90). Translating their results to our context, they proved that indeed TRI is \( \frac{1}{3} \)-integral for \( n \leq 5 \), but as \( n \) grows, vertices of TRI at fractions with arbitrarily large denominator are possible.
maximum weight stable set

identifying a...by Jebara (2009) and Sanghavi et al. (2009), which re-
treewidth 1 is a tree hence is balanced.

is either almost balanced or has treewidth 2 (a model with

for any model with block structure such that each block

Applying Corollary 13, we deduce that LP+TRI is tight


tiable settings: for each submodel, the pasted edge must in-

ncluding its special variable.

An interesting approach to MAP inference was introduced

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 Lebana (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 nand 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), hence for some potentials, the MWSS approach will fail on this model; yet LP+TRI is guaranteed to solve MAP inference efficiently for any potentials, since the treewidth is 2.

6 DISCUSSION

We have analyzed the tightness of LP relaxations on LOC

and TRI, the first two levels of the Sherali-Adams hierar-

chy, for MAP inference in binary pairwise graphical mod-

dels, demonstrating novel techniques and insights, and sig-
nificant results. The subject is of great theoretical interest

and has been studied extensively by several communities.

It is also of great practical importance given the widespread

use of LP relaxations in real-world problems. The relax-

ation on the local polytope is very popular, though recently
tighter relaxations have been implemented with impressive

results (Komodakis and Paragios, 2008; Batra et al. 2011).

We have provided intuitive proofs and derived new results that deepen our understanding and may help to provide guidance in practice, including a general decomposition re-
sult (Theorem 12). Theorem 8 on hybrid conditions (com-
bining 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 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. Fi-
nally, it provides an important step into hybrid characteriza-
tions, which remains an exciting uncharted field following success in characterizations of tractability using only topo-

logical constraints (Chandrasekaran et al., 2008), or only families of potentials (Kolmogorov et al., 2015; Thapper

and Živný 2015).

Note that by combining Theorems 8 and 12 an even larger class of models may be shown to be tight for LP+TRI by pasting almost balanced models together on edges in cer-
tain settings: for each submodel, the pasted edge must in-
clude its special variable.

In future work, we plan to examine higher order relaxations in the Sherali-Adams hierarchy, which impose consistency over larger clusters. LP+LOC=LOC3 is tight for any balanced model and we now know that LP+TRI=LOC3 is tight for any almost balanced model. It will be interesting to explore whether LP+LOC4 is tight for any model that can be rendered balanced by deleting two variables.

It may be tempting to conjecture that if LP+LOCr is tight over

a model class for some \( r \), then if an extra variable is added with arbitrary interactions, LP+LOCr+1 will be tight on the larger model. However, this is false. Consider a planar bi-

ary 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).

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 \( x_2-x_3-x_4 \) and \( x_6-x_7-x_8 \)), thus it is not always solvable by the MWSS approach. Solid blue (dashed red) edges are attractive (repulsive).

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

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+\( L_r \) 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
Acknowledgements

We thank Tomaž Slivnik and the anonymous referees for helpful comments and suggestions. 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. DS was supported by NSF CAREER award #1350965.

References


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

Tightness of LP Relaxations for Almost Balanced Models

APPENDIX: SUPPLEMENTARY MATERIAL
Tightness of LP Relaxations for Almost Balanced Models

In this Appendix, we provide the following.

- Background material:
  - § 7 Derivation of the triangle inequalities
  - § 8 Discussion of symmetry: flipping, polytope constraints and problem triangles

- The following Sections, which provide key results on the structure of weak and strong down edges, and together provide complete proofs of Theorems 8, 9 and 11 in the main paper:
  - § 9 Locking components, and 0 or 1 singleton marginals
  - § 10 Results on the Structure of Weak and Strong Down Edges in an Almost Attractive Model
  - § 11 Specification of the Perturbation for all Singleton and Edge Marginals
  - § 12 Demonstration of Consistency
  - § 13 Gathering Results and Finalizing Proofs of Theorems 8, 9 and 11

Add a high level overview of the Sections and how they fit together.

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 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:

\[
\begin{pmatrix}
q(X_i = 0, X_j = 0) & q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & q(X_i = 1, X_j = 1)
\end{pmatrix} = \begin{pmatrix}
1 + q_{ij} - q_i - q_j & q_j - q_{ij} \\
q_i - q_{ij} & q_{ij}
\end{pmatrix}
\]

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 \( \{q_i, q_j, q_k, q_{ij}, q_{ik}, q_{jk}\} \), 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 \( k = 0 \),

\[
\begin{pmatrix}
q(X_i = 0, X_j = 0) & q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & q(X_i = 1, X_j = 1)
\end{pmatrix} = \begin{pmatrix}
1 - q_i - q_j - q_k + q_{ij} + q_{ik} + q_{jk} - \alpha & q_j + \alpha - q_{ij} - q_{jk} \\
q_i + \alpha - q_{ij} - q_{ik} & q_{ij} - \alpha
\end{pmatrix}
\]

With \( k = 1 \),

\[
\begin{pmatrix}
q(X_i = 0, X_j = 0) & q(X_i = 0, X_j = 1) \\
q(X_i = 1, X_j = 0) & q(X_i = 1, X_j = 1)
\end{pmatrix} = \begin{pmatrix}
q_k + \alpha - q_{ik} - q_{jk} & q_{ij} - \alpha \\
q_{ik} - \alpha & \alpha
\end{pmatrix}
\]
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 \( \leq \) constraints with \( \leq \) 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 \( q_i + q_j \geq q_{ij} + q_{ik} \), combine the inequality from the bottom left of the upper matrix, i.e. \( q_i + \alpha - q_{ij} - q_{ik} \geq 0 \Leftrightarrow q_{ij} + q_{ik} - q_i \leq \alpha \), with the inequality from the top right of the lower matrix, i.e. \( q_{jk} - \alpha \geq 0 \Leftrightarrow \alpha \leq q_{jk} \).

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

\[
\begin{align*}
\text{Original edge marginal} & = q(X_i = 0, X_j = 0) \quad q(X_i = 0, X_j = 1) \quad q(X_i = 1, X_j = 0) \quad q(X_i = 1, X_j = 1) \\
\text{New edge marginal} & = q'(Y_i = 0, Y_j = 0) \quad q'(Y_i = 0, Y_j = 1) \quad q'(Y_i = 1, Y_j = 0) \quad q'(Y_i = 1, Y_j = 1)
\end{align*}
\]

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

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

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

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

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

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

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

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

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

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

Observe that all the triplet polytope constraints may be obtained.
We first analyze locking components, see \( q_{ij} \) for variables with 0 or 1 singleton marginals.

9.1 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 strong 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 strong 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 strong frustrated, i.e. the number of strong down edges in it is $0 \mod 2$. Edge $v_1 - v_{n-1}$ is either strong up or strong 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. 

**9.1.1 Edge marginals from locking components**

In TRI, suppose $i$ and $j$ are any two variables in a locking component, and $k$ is any other variable.

**Lemma 15.** Given $q_{ik}, q_{jk}$ is uniquely known. If one moves symmetrically, so too does the other. Specifically, if $i$ and $j$ are locked up then $q_{jk} = q_{ik}$; if $i$ and $j$ are locked down then $q_{jk} = q_k - q_{ik}$.

**Proof.** This follows by applying the TRI inequalities to the triangle $i, j, k$. We show the case where $i$ and $j$ are locked up. Let $x = q_i = q_j$. Let $y = q_{ik}$ and $r = q_{jk}$. The singleton and edge marginals are shown in Figure 8. We must show that $r = y$.

First, $q_i + q_{jk} \geq q_{ij} + q_{ik}$, i.e. $x + r \geq x + y$, hence $r \geq y$. Next, $q_j + q_{ik} \geq q_{ij} + q_{jk}$, i.e. $x + y \geq x + r$, hence $r \leq y$. 

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

This follows directly from the relevant definitions (see [4] and Lemma [4] since a problem triangle has no locking edges.

**9.2 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.
Lemma 19. If not, then the edge marginal

Proof. If variable $X_i$ has singleton marginal $q_i = 0$, then for any incident edge $(i, j)$, by the LOC constraint $q_{ij} \leq q_i$, 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


together with any variables $i$ and $j$ which have singleton marginals $q_i$ and $q_k$. Let $q_{jk}$ be the LOC-valid edge marginal for the edge $X_j - X_k$ (i.e. $q_{jk}, q_j, q_k$ satisfy (3)). It is straightforward to check that all TRI constraints (given by (9)-(10)) are satisfied. We demonstrate this for the case $q_i = 0$:

\[
\begin{align*}
q_i + q_{jk} - q_{ij} - q_{ik} &= 0 + q_{jk} - 0 - 0 = q_{jk} \geq 0 \\
q_j + q_{ik} - q_{ij} - q_{jk} &= q_j + 0 - 0 - q_{jk} = q_j - q_{jk} \geq 0 \\
q_k + q_{ij} - q_{ik} - q_{jk} &= q_k + 0 - 0 - q_{jk} = q_k - q_{jk} \geq 0 \\
q_{ij} + q_{jk} + q_{ik} - q_i - q_j - q_k + 1 &= 0 + q_{jk} + 0 - 0 - q_j - q_k + 1 = q_{jk} - (q_j + q_k - 1) \geq 0
\end{align*}
\]

\[\square\]

10 Results on the Structure of Weak and Strong Down Edges in an Almost Attractive Model

Throughout this Section, we assume an almost attractive model, where edge marginals have been optimized over TRI given singleton marginals. Further, as justified by (9) we assume no locking edges or variables that have singleton marginal 0 or 1.

Lemma 17. In every triplet of variables, at most one triplet constraint is tight.

Proof. If any two triplet constraints hold, it is easily seen that this implies a locking edge. We show one case: suppose $a + bc = ab + ac$ and $b + ac = ab + bc$, then adding equations gives $a + b = 2ab$, but by a LOC constraint, $ab \leq \min(a, b)$, hence we must have $ab = a = b$, i.e. a strong up locking edge.

\[\square\]

Lemma 18. Any weak edge $uv$ must be tight in some triplet constraint, that is there must exist some variable $w$ s.t. there is a tight triplet constraint in $u, v, w$.

Proof. If not, then the edge marginal $uv$ may be perturbed up and down by a sufficiently small $\epsilon$ without violating any LOC or TRI constraints, hence we cannot be at a vertex.

\[\square\]

Lemma 19 (When 2 strong edges in a triangle force the 3rd edge to be strong). Consider triangle $abc$ where edges $ab$ and $ac$ are strong. The following cases force the edge $bc$ (all cases may be regarded as flippings of the first case):

(i) $ab, ac$ up and $a \in [b, c]$ (a is in the middle) $\Rightarrow$ $bc = \min(b, c)$ is strong up;
(ii) $ab, ac$ down where one is 0 and the other is $> 0$ $\Rightarrow$ $bc$ is strong up (with marginal equal to the end of the 0 edge from $a$);
(iii) $ab$ up with $a > b$, and $ac$ down with $ac = 0$ $\Rightarrow$ $bc = 0$ strong down;
(iv) $ab$ up with $a < b$, and $ac$ down with $ac > 0$ $\Rightarrow$ $bc = b + c - 1$ strong down.

Proof. These are easily shown by applying TRI constraints to $abc$. We demonstrate the first case by applying the inequality $a + bc \geq ab + ac$: if $b \leq a \leq c$ then the inequality is $a + bc \geq b + a \Rightarrow bc = b$; similarly, $c \leq a \leq b \Rightarrow bc = c$.

\[\square\]
Adrian Weller, Mark Rowland, David Sontag

Figure 9: An illustration of the situation considered in Lemma 21. If $wx \rightarrow xz \rightarrow xy$ is a thistle, then so too is $wx \rightarrow xy$. Broken wavy edges indicate edges which are either strong down or weak (but not strong up), i.e. they are dw edges.

Figure 10: An illustration of the 2 structures which cannot occur in an almost attractive model if edge marginals are optimized; see Theorem 23. Solid blue edges are strong up, wavy red edges are strong down, and dashed green edges are weak. The right structure is equivalent to the left by a flipping of $s$.

Notation: We say that any edge which is strong down or weak is a dw edge. Thus, any edge which is not strong up is dw.

In an almost attractive model, any dw edge $xy$ not incident to $s$ must be being held down by some TRI constraint, say in triplet $wxy$. This must have one of two forms, either (i) $x + wy = wx + xy$, or (ii) $y + wx = wy + xy$. (The other 2 TRI inequalities, if tight, would hold up $xy$.) In case (i), we say that $wy$ is holding down $xy$ and write $wy \rightarrow xy$. In case (ii), $wx$ is holding down $xy$ and we write $wx \rightarrow xy$.

Note that $wy \rightarrow xy$ is equivalent to $wy \rightarrow wx$; both mean that $x + wy = wx + xy$.

Definition 20. A thistle from edge $e_1$ to edge $e_k$ of length $k$ is a sequence of edges $e_1 \rightarrow e_2 \rightarrow \cdots \rightarrow e_k$ where there is one variable in common between successive edges, that is $|\{e_i, e_{i+1}\} \cap \{e_{i+1}, e_{i+2}\}| = 1$ and each edge is holding down the next for all $i = 1, \ldots, k-1$.

An example thistle might be of the form $uw \rightarrow vw \rightarrow wx \rightarrow xy$. Note though that in general, a thistle may not be a direct path. For example, a thistle could take the form $uw \rightarrow vw \rightarrow vx \rightarrow xy$. In this example, we think of $vw$ as a ‘thorn’ that sticks out to the side, which is why we call these structures thistles. We next provide two Lemmas which show that thistles of length 3 can be ‘contracted’ to length 2.

Lemma 21. If $xw \rightarrow xz \rightarrow xy$ is a thistle (note this has a thorn), then so too is $xw \rightarrow xy$.

Proof. Consider Figure 9. We know that $xw$ is holding down $xz$ and $xz$ is holding down $xy$. Further we have an inequality for triangle $wxy$. Hence we have

\begin{align*}
z + xw &= xz + wz & (12) \\
y + xz &= xy + yz & (13) \\
y + xw &\geq wy + xy & (14)
\end{align*}

Now (12) + (13) gives $xw = xy + yz + wz - y - z$. Substituting into (14) gives $yz + wz \geq z + wy$. But now observe that we have $z + wy \geq wz + yz$ as a triplet constraint in $wyz$, hence (14) must hold with equality, which proves the result.

Lemma 22. If $wx \rightarrow xy \rightarrow yz$ is a thistle (note this follows a path with no thorn), then so too is $wz \rightarrow yz$.
Proof. The proof is similar to that of Lemma 21. We have that \( wx \) is holding down \( xy \), and \( xy \) is holding down \( yz \). Further, we use an inequality for the triangle \( wyz \):

\[
\begin{align*}
y + wx &= wy + xy & \text{(15)} \\
z + xy &= xz + yz & \text{(16)} \\
y + wz &\geq wy + yz & \text{(17)}
\end{align*}
\]

Now \((15) + (16)\) yields \( yz = y + z + wx - wy - xz \). Substituting into \((17)\) and rearranging gives \( wz + xz \geq z + wx \). But we have the TRI inequality \( z + wx \geq wz + xz \), so equality must be attained in \( wz + xz \geq z + wx \), and so we must have equality in \((17)\) which yields the result. \(\square\)

Notice that in both Lemmas 21 and 22, the \( w \) variable in the first edge features exactly once in the conditions of the Lemmas, and then again features as one of the ends of the edge holding down the other in the conclusion of the result.

Using these earlier Lemmas, we show the following key structural result on \( dw \) edges.

**Theorem 23** (\( dw \) edges away from \( s \)). *Every \( dw \) edge \( xy \) which is not incident to \( s \) is pulled down by an edge incident to \( s \), i.e. either \( sx \rightarrow xy \) or \( sy \rightarrow xy \).*

Proof. Any \( dw \) edge \( xy \) not incident to \( s \) is attractive, hence must be held down by another edge (i.e. \( xy \) must be in a triplet where there is a binding TRI constraint which upper bounds \( xy \)), which WLOG we may assume is \( ux \) for some \( u \). If \( u = s \) then we are done. Otherwise \( wx \) is attractive, and must be \( dw \) (since if \( wx \) were strong up, it is easily checked that it could not hold down \( xy \), i.e. \( y + wx \geq uy + xy \) will always hold, even if \( xy \) is strong up) and we may keep repeating the argument to grow a thistle back from \( xy ; . . . \rightarrow ux \rightarrow xy \). As we work back, since the graph is finite, one of the following two cases must occur:

1. We eventually hit an edge incident to \( s \). The result then follows by repeatedly applying Lemmas 21 or 22.
2. We have a sub-thistle, the edges of which form a chordless cycle in the graph of length \( k \geq 3 \), \( a_1a_2 \rightarrow a_2a_3 \rightarrow \cdots a_ka_1 \). Now repeatedly apply Lemmas 21 or 22 alternately to the sub-thistle given by the first three edges until we obtain either: \( a_1a_k \rightarrow a_{k-1}a_k \rightarrow a_ka_1 \) or \( a_1a_{k-1} \rightarrow a_{k-1}a_k \rightarrow a_ka_1 \). In either case, this implies two tight triangle inequalities in \( a_1a_{k-1}a_k \) (this follows directly from the definition above of the \( \rightarrow \) notation; for example, \( a_1a_k \rightarrow a_{k-1}a_k \rightarrow a_ka_1 \)) means \( a_{k-1} + a_1a_k = a_1a_{k-1} + a_{k-1}a_k \) (from \( a_1a_k \rightarrow a_{k-1}a_k \)) and also \( a_1 + a_{k-1}a_k = a_1a_{k-1} + a_1a_k \) (from \( a_{k-1}a_k \rightarrow a_ka_1 \)), which is a contradiction by Lemma 17.

Note that as a consequence of this Theorem, the two configurations shown in Figure 10 cannot occur. \(\square\)

We show a strengthening of the result if the \( dw \) edge is strong down.

**Lemma 24** (Strong down edges away from \( s \)). *If \( xy = 0 \) is a strong down edge with \( s \notin \{x, y\} \), then either: \( sx = x \) is strong up and \( sy = 0 \) is strong down; or \( sx = 0 \) is strong down and \( sy = y \) is strong up.*

If \( xy > 0 \) is a strong down edge with \( s \notin \{x, y\} \), then either: \( sx = x \) is strong up and \( sy > 0 \) is strong down; or \( sx > 0 \) is strong down and \( sy = s \) is strong up.

Proof. By Theorem 23, we have \( sx \rightarrow xy \) or \( sy \rightarrow xy \). The remainder of the statement of the proof follows as a straightforward application of the relevant TRI constraint. We show the case \( xy = 0 \) and \( sx \rightarrow xy \): We have \( y + sx = sy + xy = sy \). Rewrite this as \( y - sy + sx = 0 \). Both terms are \( \geq 0 \) hence must both be exactly zero. \(\square\)

### 11 Specification of Complete Symmetric Perturbation (including all edges)

Throughout this Section, we assume an almost attractive model with special variable \( s \), where edge marginals have been optimized over TRI given singleton marginals. Further, as justified by \( 9 \) we assume no locking edges or variables that have singleton marginal 0 or 1.

We shall specify a perturbation for all singleton and all edge marginals with a number which is -1, 0 or 1 for each marginal. The perturbation up is formed by taking the vector of all these numbers and multiplying by a small \( \epsilon \). The perturbation
down is exactly the negative of the perturbation up. \( \epsilon \) is to be chosen sufficiently small s.t. any constraint (this includes all TRI constraints, all LOC constraints, and all constraints on a marginal being \( \geq 0 \) and \( \leq 1 \)) which was not tight initially, remains so after either perturbation. In order for both perturbations to remain in TRI, we shall demonstrate that all tight TRI constraints (and also all LOC constraints, see [11.2.1] are exactly maintained in all cases.

11.1 Rule for Singleton Marginals

The perturbation for the singleton marginal of the variable \( s \) is 0. For any other variable \( v \in V \setminus \{s\} \), its perturbation depends on its edge marginal to \( s \), i.e. \( sv \), according to the following exhaustive options (recall that we are assuming no locking edges):

\[
\begin{align*}
\text{\( v \) moves by } +1 & \text{ if \( v \) is strong up to \( s \) and } v > s, \text{ or } v \text{ is strong down to } s \text{ and } v + s < 1, \\
\text{\( v \) moves by } -1 & \text{ if \( v \) is strong up to } s \text{ and } v < s, \text{ or } v \text{ is strong down to } s \text{ and } v + s > 1, \\
\text{\( v \) moves by } 0 & \text{ if } v \text{ has a weak edge to } s.
\end{align*}
\]  

We remark that this perturbation has the appealing property that it maps to itself (actually it maps to the negative of itself, but that is equivalent since we perturb up and down) under a flipping of \( s \) (if a perturbation works for all almost attractive models, then the version obtained from it by flipping \( s \) must also work for all almost attractive models, since flipping \( s \) is a bijection from the set of all almost attractive models to itself).

11.2 Rule for Edge Marginals

Given the changes in (18) for singleton marginals, we now show the perturbation for edge marginals.

11.2.1 Strong Edges

If an edge is strong (i.e. a LOC constraint is tight), we may immediately determine the perturbation required in order that LOC constraints are respected for both perturbations up and down. Specifically:

\[
\begin{align*}
uv \text{ moves with } \min(u, v) & \text{ if } uv \text{ is strong up} \\
uv \text{ moves with } \max(0, u + v - 1) & \text{ if } uv \text{ is strong down.}
\end{align*}
\]  

11.2.2 Note on Consistency, Remaining within TRI

The above rules clearly ensure that our perturbed marginals remain in LOC. Note that for any edge that had a tight LOC constraint, i.e. was strong, the above rules exactly maintain this constraint when perturbed. We adopt this idea to ensure that we shall also remain in TRI. That is, for our perturbed marginals to be in TRI, it is clearly sufficient if we ensure that every TRI constraint that was tight, is exactly maintained for the perturbed marginals. In order to demonstrate that our perturbation satisfies this condition, we shall explicitly prescribe all perturbations for all weak edges, and show that our prescribed perturbation exactly maintains all TRI constraints that are tight. In order to do this, we shall have to demonstrate that our prescribed changes for edges are consistent with the change that is necessary in all other triplets to preserve tight TRI constraints. This is what we mean by consistency, which we explore fully in [12].

11.2.3 Weak Edges Incident to \( s \)

The perturbation for a weak edge \( sw \) incident to \( s \) is -1. This is chosen since it is necessary to ensure consistency for any TRI constraint involving the weak edge and any 2 strong edges, as we show in [12.2.1].

Note that we have now specified all edges (weak and strong) incident to \( s \).

11.2.4 Weak Edges Not Incident to \( s \), \( \delta \) Notation

Given the above specifications, we may now use Theorem [23] to prescribe the change necessary for any weak edge \( uv \) not incident to \( s \) in order to maintain consistency. We adopt the notation \( \delta(v) \in \{-1, 0, +1\} \) for the perturbation of a singleton marginal \( v \), and \( \delta(uv) \in \{-1, 0, +1\} \) for the perturbation of an edge marginal \( uv \). There are exactly 7 possible cases to consider. In each case, it is straightforward to compute the required perturbation for the weak edge not incident to \( s \), as shown in Figure [11]. We provide more detail below.
If a weak edge $uv$ is not incident to $s$, then by Theorem 23, it lies in a tight triangle with $s$, and as described in §11.2.4, we may deduce its necessary edge perturbation by considering the tight TRI constraint in the triangle $suv$. The 7 possible cases are shown in Figure 11(a)-(g).

Note that by Lemma 19, the two configurations shown in Figure 12 cannot have a tight TRI constraint without contradicting the weakness of $uv$. Thus, these are omitted from Figure 11 and may be excluded from further analysis. Observe that a configuration of the form given in Figure 12 may be obtained by flipping the variable $s$ in Figure 11d, and the configurations shown in Figures 11c and 11e may similarly be obtained from those in Figures 11b and 11c by flipping $s$. We may therefore exclude these cases from our analysis too, and need only show here that the perturbations defined for the weak edges in Figures 11a, 11b, 11c, and 11d are consistent.

The perturbations for the weak edge $uv$ that are indicated in the various configurations of Figure 11 may be derived straightforwardly by considering the tight TRI constraint in each case, using the prescribed perturbation for the other edges as given by $\{11.1\}$ and $\{11.2.1\}$ and observing what perturbation of the weak edge is implied in order to maintain tightness of the relevant TRI constraint. We go through cases:

- In Figure 11a, the tight TRI constraint must be either $u + sv = su + uv$ or $v + su = uv + sv$. In either case, by noting that $\delta(u) = \delta(v) = 0$ and $\delta(sv) = \delta(sv) = -1$, as prescribed in §11.2.3, it follows that to maintain tightness of the TRI constraint, we must have $\delta(uv) = 0$.

- In Figure 11b, the tight TRI constraint must be $u + sv = su + uv$. Noting that $\delta(u) = -1$, $\delta(sv) = -1$, and $\delta(sv) = \delta(u) = -1$, we must have $\delta(uv) = -1$.

- In Figure 11c, the tight TRI constraint must be $u + sv = su + sv$. Noting that $\delta(u) = 1$, $\delta(sv) = -1$ and $\delta(sv) = \delta(s) = 0$, we must have $\delta(uv) = 0$.

- In Figure 11d, the tight TRI constraint must be $v + su = sv + uv$. Noting that $\delta(v) = 1$, $\delta(sv) = 0$, and $\delta(sv) = \delta(s) = 0$, we must have $\delta(uv) = 1$. 

---

Figure 11: Cases where a weak edge $uv$ is not incident to $s$. By Theorem 23, there must be a tight TRI constraint in $suv$. Here we show the possible forms with the implied perturbation for the weak edge. Note that the forms in the lower row may each be obtained from an appropriate form in the upper row by flipping $s$, so need not be considered separately. Specifically, under flipping $s$ we have $a \leftrightarrow a, b \leftrightarrow e, c \leftrightarrow f, d \leftrightarrow g$.

Figure 12: Cases which are not possible when $suv$ has a tight TRI constraint since each implies that $uv$ is strong down.
12 Demonstrating Consistency

We shall show that the perturbation prescribed in §11.2 maintains all tight TRI constraints, which is sufficient for us to stay within TRI after perturbing both up and down.

We must consider all cases of a triplet with a tight TRI constraint. We divide the cases up into 4 exhaustive classes:

(i) The triplet contains 0 weak edges (hence 3 strong edges), we call this 0-wedge consistency. See §12.1
(ii) The triplet contains 1 weak edge (hence 2 strong edges), we call this 1-wedge consistency. See §12.2
(iii) The triplet contains 2 weak edges (hence 1 strong edge), we call this 2-wedge consistency. See §12.3
(iv) The triplet contains 3 weak edges (hence 0 strong edges), we call this 3-wedge consistency. See §12.4

12.1 0-wedge consistency

In this Section we consider a triangle with 3 strong edges. Recall that by construction, our perturbation maintains the nature of all strong edges (strong up stay strong up, strong down stay strong down). We make the following observation.

Lemma 25. In a triangle with three strong edges including an even number of strong down edges (so the triangle is not strong frustrated), all TRI constraints are always satisfied.

Proof. This follows by straightforward checking of the TRI constraints §9-10. We demonstrate one case. Suppose $abc$ is a triangle with 3 strong up edges. We shall show that $a + bc \geq ab + ac$. Consider $f = a + bc - ab - ac$, we shall show $f \geq 0$. We have $f = a + \min(b, c) - \min(a, b) - \min(a, c)$, clearly symmetric in $b$ and $c$, thus we may consider just 3 subcases:

\[
\begin{align*}
& a \leq b \leq c \quad \Rightarrow \quad f = a + b - a - a = b - a \geq 0 \\
& b \leq a \leq c \quad \Rightarrow \quad f = a + b - b - a = 0 \\
& b \leq c \leq a \quad \Rightarrow \quad f = a + b - c = a - c \geq 0.
\end{align*}
\]

Hence we need consider only triangles that are strong frustrated. We may rule out 3 strong down edges.

Lemma 26. A triangle with 3 strong down edges cannot occur.

Proof. Lemma §24 shows that $s$ cannot be in such a triangle. Now applying Lemma §24 to each edge in turn around the triangle yields a contradiction: we must alternate between strong up and strong down edges to $s$, yet this is not possible since we have an odd number of edges (if an edge is both strong up and strong down, one end must have marginal of 0 or 1, which we are assuming cannot occur).

Thus we need consider only the case of a strong frustrated triangle $abc$ that has 1 strong down edge $bc$ and 2 strong up edges $ab, ac$, where a TRI constraint is tight. By Lemma §19 we must have $a < b, c$ or $a > b, c$. It is easily checked that the only TRI constraint of concern is where $a + bc = ab + ac$, hence we may assume that this holds. These are called problem triangles of type (i) and (ii) in the main paper §4. Note that $s$ could be $b$ or $c$ (in which case, we assume $b$ WLOG) but not $a$ by Lemma §24. See Figure §13 for illustrations of the four possibilities.

Considering first the cases where $s$ is in the triangle. If $a > s, c$ then we have $\delta(a) = +1, \delta(sc) = 0, \delta(sa) = \delta(s) = 0, \delta(ac) = \delta(c) = +1 \Rightarrow \delta(a + sc - sa - ac) = 1 + 0 - 0 - 1 = 0$ so we have consistency. If $a < s, c$ then $\delta(a) = -1, \delta(sc) = \delta(c) = -1, \delta(sa) = \delta(a) = -1, \delta(ac) = \delta(a) = -1 \Rightarrow \delta(a + sc - sa - ac) = -1 - 1 + 1 + 1 = 0$ as required.

If $s$ is not in the triangle then we may use Lemma §24 to give the edges from $s$ to $b$ and $c$, which determine their perturbations. WLOG we shall assume that $sb$ is strong down and $sc$ is strong up. It remains to determine the perturbation change to $a$, which we shall do by considering the edge $sa$.

If $a > s, c$ (case $c$ in Figure §13) then we have $sb = 0, sc = c, ac = c$. Also $a + bc = ab + ac \Rightarrow a = b + c$. From triangle $sca$ we have $c + sa \geq sc + ac \Rightarrow sa \geq c + c - c = c$ while from $sba$ we have $a + bs \geq ab + sa \Rightarrow sa \leq a + 0 - b = c$. Hence $sa = c$ a weak edge. Now $\delta(a + bc - ab - ac) = 0 + 0 - 1 = 1 = 0$ as required.
Tightness of LP Relaxations for Almost Balanced Models

Figure 13: The four possible types of triangles $abc$ with one strong down edge $bc$ and two strong up edges to consider for $0$-wedge consistency. We have $a + bc = ab + ac$, either $a > bc$ with $bc = 0$, or $a < b, c$ with $bc > 0$. On the left we have the cases where $s$ is in the triangle. See §12.1.

Figure 14: The two possible types of triangles with one weak edge incident to $s$

If $a < s, c$ (case d in Figure 13) then we have $sb > 0$, $sc = s$, $ac = ab = a$. Also $a + bc = ab + ac \Rightarrow a = b + c - 1$. Applying the same TRI inequalities as in the last case, we obtain $sa = s + b - 1$, again a weak edge. Now $\delta(a + bc - ab - ac) = 0 + 0 - 0 + 0 = 0$ as required.

12.2 1-wedge consistency

We split the 1-wedge class into subclasses. We first consider in §12.2.1 the case that the 1 weak edge is incident to $s$. Then in the following Sections, we demonstrate consistency exhaustively for all possible configurations of weak edges that are not incident to $s$. These are illustrated in Figure 11. We need consider only cases shown in 11a to 11d since the remaining cases may be obtained from these by flipping $s$.

12.2.1 Perturbation of a weak edge incident to $s$ consistent with a TRI including 2 strong edges

As in §11.2.3 let $sw$ be a weak edge incident to $s$. Recall that we prescribed $\delta(sw) = -1$. Here we shall consider any possible triangle involving a third variable $v$ with $sw$ and $vw$ strong, and demonstrate consistency.

By Lemma 24, $vw$ cannot be a strong down edge, hence $vw$ must be strong up. There are therefore two cases to consider: (i) both $sv$ and $vw$ are strong up; and (ii) $sv$ is strong down, and $vw$ is strong up. See Figure 14.

We first consider case (ii). By Lemma 19, because $sw$ is weak, we must have one of the 2 subcases shown in Figure 15. The only possible tight TRI constraint, which must therefore apply, is $w + sv = sw + vw$.

In order to maintain this TRI constraint through the perturbations, we must have $\delta(w) + \delta(sv) = \delta(sw) + \delta(vw)$.

Using our rules for perturbation from §11.1 and §11.2.1, in both subcases this gives $\delta(sw) = -1$ which is consistent with our rule in §11.2.3.

We now consider case (i) where $sv$ and $sw$ are both strong up. The only possible tight TRI constraint, which must hold, is $v + sw = sv + vw$. By Lemma 19, we must have either $v < s, w$ or $v > s, w$, see Figure 16. In either case, to preserve the
tightness of the TRI constraint, following our rules for perturbation from [11.1] and [11.2.1] we must have \( \delta(sw) = -1 \), which is consistent with our rule in [11.2.3].

### 12.2.2 1-wedge Consistency of Weak Edge Perturbations Defined in Figure 11a

Here we prove that for any weak edge \( uv \) of the form appearing in Figure 11a if there exists another variable \( x \) such that \( uvx \) is a triangle with a tight TRI constraint, and \( ux, vx \) are strong edges, then the tight TRI constraint is maintained the prescribed perturbation for \( uv \). This scenario is illustrated in Figure 17.

First note that by Lemma 24, \( ux \) and \( vx \) cannot be strong down. Therefore the only scenario to consider in this case is when \( ux \) and \( vx \) are strong up. Recall from Lemma 19, we must have either \( x < u, v \) or \( x > u, v \). Note also that we have the prescribed perturbations \( \delta(u) = \delta(v) = \delta(uv) = 0 \). Since the only possible tight TRI constraint in \( uvx \) that doesn’t contradict the weakness of \( uv \) is \( x + uv = ux + vx \), this equality must hold. By considering Figure 18, note that in each case, we must prove that \( \delta(x) = 0 \) in order for tightness of this constraint to be maintained. Thus, it is sufficient in each case to prove that \( sx \) is weak.

First, we consider \( x > u, v \) - see Figure 18a. In the tight triangle \( suv \), one of the TRI constraints \( u + sv \geq su + uv \) and \( v + su \geq sv + wv \) must be tight; without loss of generality, we take \( u + sv = su + uv \). In the tight triangle \( uvx \), it must be the case that the tight TRI constraint is \( x + uv = ux + vx \). From these two equations, we obtain \( x = v - sv + su \). Now considering TRI inequalities in \( svx \), we note \( v + sx \geq sv + vx \), so \( v + sx \geq sv + v \), and so \( sx \geq sv \). We also have \( x + sv \geq sx + vx \), which leads to \( su \geq sx \) (by using the fact that \( x = v + su - sv \)). So we obtain

\[
\min(s, x) \geq \min(s, u) > su \geq sx \geq sv > 0
\]

Therefore if we can show that \( sx \neq s + x - 1 \), we have that \( sx \) is weak, so that \( \delta(x) = 0 \), and so the tight TRI constraint \( x + uv = ux + vx \) is maintained under the perturbation, as we set out to show. To show this, suppose \( sx = s + x - 1 \), and consider the TRI constraint \( u + sx \geq su + ux \). Substituting in our expression for \( x \), we obtain \( s + v - 1 \geq sv \), contradicting weakness of \( sv \). Therefore \( sx \) is weak, and \( \delta(x) = 0 \), as required.

Next, we consider \( x < u, v \), as in Figure 18b. Again, for the tight TRI constraint in \( suv \) we may assume without loss of generality that \( u + sv = su + uv \). The only TRI constraint that can be tight in \( uvx \) (without contradicting weakness of \( uv \)) is \( x + uv = ux + vx \), which implies \( uv = x \), so \( x = u + sv - su \). Considering the TRI constraint \( u + sx \geq su + ux \) gives \( sx \geq sv \), and considering the TRI constraint \( x + sv \geq sx + vx \) gives \( sv \geq sx \). Therefore we have \( sv = sx \), and so immediately we have \( sx > 0 \) and \( sx < s \). We now just need to rule out \( sx = s + x - 1 \) and \( sx = x \). If \( sx = s + x - 1 \), then by considering the TRI constraint \( v + sx \geq vx + sv \), we obtain \( sv \leq s + v - 1 \), contradicting weakness of \( sv \). If \( sx = x \), then we obtain \( su = u \) is strong up, a contradiction.
12.2.3 1-wedge Consistency of Weak Edge Perturbations Defined in Figure 11b

Here we prove that for any weak edge $uv$ of the form appearing in Figure 11b, if there exists another variable $x$ such that $uxv$ is a triangle with a tight TRI constraint, and $ux$, $vx$ are strong edges, then the tight TRI constraint is maintained the prescribed perturbation for $uv$. This scenario is illustrated in Figure 19.

First, we note that $xv$ cannot be strong down, by Lemma 24. Therefore we take $xv$ strong up. Note also that since $u < s$, if $ux$ is strong down, then it has edge marginal $0$ and $sx$ is strong down with edge marginal $0$ too. Recall also that if $ux$ is strong up, then by Lemma 19 we have $x > u, v$ or $x < u, v$. Figure 20 illustrates these cases.

In Figure 20a, note that the tight TRI constraint in $uxv$ must be $v + ux = uv + vx$. Noting that in this case, we have $\delta(v) = 0$, $\delta(ux) = 0$, $\delta(uv) = -1$, if $v > x$, then $\delta(vx) = \delta(x) = -1$ (so the tightness of the TRI constraint is maintained). If $v < x$, then the TRI constraint $v + sx \geq xv + sv$, implies $sv = 0$, a contradiction.

In Figure 20b, note that the only possible tight TRI constraint in $uxv$ is $x + uv = ux + vx$ (all others contradict weakness of $uv$). Note also that $u + sv = su + uv$ is the only possible tight TRI constraint in $swv$, so $sv = uv$. Lastly, we have the TRI constraint $x + sv \geq sx + vx$. But $x + sv = u + v$, so $u \geq sx$. But considering the TRI constraint $u + sx \geq su + ux$ gives $sx \geq u$. So $sx = u$ is weak. So we have $\delta(x) = 0$, $\delta(u) = -1$, $\delta(v) = 0$, and $\delta(uv) = -1$, so the tightness of $x + uv = ux + vx$ is maintained.

In Figure 20c, note that by considering the TRI constraint $u + sx \geq ux + su$, we obtain $sx = x$. We then note that the tight TRI constraint in $uxv$ must be $x + uv = ux + vx$ (all others contradict the weakness of $uv$). Note that we have $\delta(x) = -1$, $\delta(uv) = -1$, $\delta(ux) = -1$, and $\delta(vx) = -1$, so the tightness of the TRI constraint is maintained.
12.2.4 1-wedge Consistency of Weak Edge Perturbations Defined in Figure 11c

Here we prove that for any weak edge $uv$ of the form appearing in Figure 11c, if there exists another variable $x$ such that $uxv$ is a triangle with a tight TRI constraint, and $ux, vx$ are strong edges, then the tight TRI constraint is maintained by the prescribed perturbation for $uv$. This scenario is illustrated in Figure 21.

First, we note that $xv$ cannot be strong down, by Lemma 24. Therefore we take $xv$ strong up. Note also that since $u < s$, if $ux$ is strong down, then it has edge marginal $u + x - 1$ and $sx$ is strong down with edge marginal 0 too. Recall also that if $ux$ is strong up, then by Lemma 19 we have $x > u, v$ or $x < u, v$. Figure 22 illustrates these cases.

In Figure 22a note that if $vx = x$, then by considering the TRI constraint $v + sx \geq su + xv$ implies that $sv$ is strong down, a contradiction. So $vx = v$. Note that the only possible tight TRI constraint in $uxv$ is $v + ux = uv + vx$, and we have $\delta(v) = 0, \delta(uv) = 0, \delta(vx) = 0$, and $\delta(ux) = 0$, so the TRI constraint remains tight.

In Figure 22b note that from the TRI constraint $u + sx \geq su + ux$, we obtain $sx \geq s$, and so $sx = s$. The only possible tight TRI constraint in $uxv$ is $x + uv = ux + vx$ (all others contradict the weakness of $uv$). But then note we have $\delta(x) = 1$, $\delta(ux) = 1, \delta(vx) = 0$ and $\delta(uv) = 0$, so the TRI constraint remains tight.

In Figure 22c note that the only possible tight TRI constraint in $uxv$ is $x + uv = ux + vx$ (all others contradict the weakness of $uv$), so we obtain $uv = x$. Note that $u + sx \geq su + ux$, so $sx \geq s + x - u$. Also, $x + sv \geq sx + vx$, so $sv \geq sx$. But the only possible tight TRI constraint in $suw$ is $u + su = sv + uv$, and this yields $s + x - u \geq sx$, so $sx = s + x - u$. Note this quantity is less than $s$ and $x$, so $sx$ not strong up; it is greater than $s + x - 1$, and if it is equal to 0, then we have $sv = 0$, a contradiction. Therefore $sx$ is weak. From this, note that $\delta(x) = 0, \delta(ux) = 0, \delta(vx) = 0$ and $\delta(uv) = 0$, so the TRI constraint in $uxv$ remains tight.

12.2.5 1-wedge Consistency of Weak Edge Perturbations Defined in Figure 11d

In this section, we prove that for any weak edge $uv$ of the form appearing in Figure 11d, if there exists another variable $x$ such that $uxv$ is a triangle with a tight TRI constraint, and $ux, vx$ are strong edges, then the tight TRI constraint is maintained by the prescribed perturbation for $uv$. This scenario is illustrated in Figure 23.

There are three separate realisations of the scenario in Figure 23 to consider; see Figure 24.

Firstly, the case where $ux$ is strong down - this is illustrated in Figure 24a. Since $ux$ is incident to $us$, which has edge marginal 0, $ux = 0$ too, by Lemma 24. Again by Lemma 24 $sx$ is strong up and equal to $x$. Applying the TRI constraint $s + xv \geq sx + sv$ yields $vx \geq x$, so $vx = x$ and is strong up. Finally, checking which TRI constraint can hold in the tight
triangle $uvx$, we note that the only possibility not contradicting the weakness of $uv$ is $v+ux = uv + vx$, which leads to $s = x$, so we have a locking component and need not consider this example further.

Secondly, we consider $ux, vx$ strong up; see Figures 24b-24c. Recall from Lemma 19 that we need only consider $x > u, v$. First consider $x < u, v$. The only TRI constraint in $uvw$ that can be tight without contradicting the weakness of $uv$ is $x + uv = xu + xv$, so $uv = x$. But from the tight TRI constraint in $usv$, we get $uv = v - s$, so $x = v - s$. Now considering the TRI constraint $v + sx \geq vs + vx$ gives $sx = 0$. Therefore we have $\delta(x) = 1, \delta(ux) = 1, \delta(xv) = 1, \delta(xu) = 1$, and verify that the TRI constraint remains tight under the perturbation. If $x > u, v$, then considering $v + sx \geq vs + vx$ gives $sx = s$, so again we obtain $\delta(x) = 1, \delta(ux) = 1, \delta(xv) = 1, \delta(xu) = 1$, and verify that the TRI constraint remains tight under the perturbation.

Finally, we consider $ux$ strong up and $vx$ strong down; see Figure 24d. $vx$ strong down implies that $sx$ strong down, and $sv = s$ implies that both $sx$ and $vx$ are strong down with edge marginal greater than 0. The only TRI constraint in $uxv$ that can be tight without contradicting the weakness of $uv$ is $u + vx = uv + vx$. This implies $ux = x + (u + v - 1 - uv) < x$ (as $uv$ not strong down), so $ux = u$, and so $uv = x + v - 1$. But note since $v + su = sv + uv$, we have $uv = v - s$, and so $s = 1 - x$. Thus we have a locking component and need not consider this case further.

12.3 2-wedge consistency

12.3.1 The case where the 2 weak edges are both incident to $s$

This case is shown in Figure 25. The two possible tight TRI constraints are $s + sv = uv + su$ or $v + su = uv + sv$. In either case, it is clear that we obtain a consistent conclusion that $\delta(uv) = 0$ (consistent with $\delta(u) = \delta(v) = 0$ and hence the strong edge $uv$ does not move).
12.3.2 All other cases of 2-wedge consistency

For all these cases, we consider a tight triangle \( xyz \) away from \( s \), with \( xy, yz \) weak and \( xz \) strong. We note that by earlier arguments, \( sxy \) and \( syz \) must be triangles from Figure 11. Therefore it is sufficient to consider all pairs of triangles \( sxy \) and \( syz \) from Figure 11 and show that the tight TRI constraint in \( xyz \) remains tight under the perturbation. A priori this gives \( 7 \times 7 = 49 \) cases to check. However, by flipping \( s \) if necessary, \( sxy \) may always be taken to be one of a)-d) from Figure 11, reducing the burden to 28 cases. We further note that by symmetry we always take \( sxy \) to be a triangle listed no later in Figure 11 than triangle \( syz \) - this rules out a further 6 cases to check. The remaining cases are exhaustively examined below.

Note that the nature of the edge \( xz \) is not specified explicitly by the triangles \( sxy \) and \( syz \). However, since in this section we consider triangles \( xyz \) with exactly two weak edges, we do not consider the cases where \( xz \) is weak - these are covered in \( \S \) 12.4.

In some cases, we will want to argue that certain combinations of tight TRI constraints and strong edges contradict our assumptions that we have no locking edges, and/or our assumptions about which edges are weak. It is possible, but laborious, to prove these contradictions of our assumptions algebraically: here we briefly explain a MATLAB script written to verify these contradictions automatically, in the context of its use in \( \S \) 12.3.4. In this case, we wish to show that it cannot be the case that \( sz = z, xz = z \), all other edges weak, and \( z + sy = sz + yz, z + xy = xz + yz \) and \( y + sx = sy + xy \) without our assumptions of no locking edges, or the weakness of the other edges, being contradicted. To do this, we run the script shown at the top of Listing 1.

Listing 1: Example script used in this section

```plaintext
>> equalities = {'z=sz','z=xz','z+sy','z+xy','y+sx'};

% Test weak edges:
testWeakness(equalities, 'sx')
testWeakness(equalities, 'sy')
testWeakness(equalities, 'xy')
testWeakness(equalities, 'yz')
% Test whether strong edges lock:
testLocking(equalities, 'sz', 'up')
testLocking(equalities, 'sx', 'up')
```

The variables \( \text{equalities} \) is a cell containing strings, which code for which LOC and TRI constraints we would like to take to be tight. This gives rise to a new polytope, the restricted polytope given by intersecting TRI with all of these constraints. The function \( \text{testWeakness} \) examines a particular input edge \( uv \) in the graph to see whether it is always strong. This is implemented by checking whether any of the equations \( uv = 0, uv = u + v - 1, uv = u, uv = v \) always hold in the restricted polytope. All four equations are checked in a similar way; for example, to check whether \( uv = u \) at all points in the restricted polytope, two linear programs are set up to maximise and minimise the quantity \( uv - u \) over the restricted polytope. If the maximum and minimum are both found to be 0 (in practice, we use a threshold of \( 1e-6 \)), then we deduce that \( uv = u \) at all points in the polytope, and so we deduce that edge \( uv \) is forced to be strong, given the set of constraints assumed in the \( \text{equalities} \) variable. Similarly, the function \( \text{testLocking} \) checks whether a particular edge is locking up or down, by checking whether the two incident edge marginals are always equal (in the case of locking up) or always sum to 1 (in the case of locking down) at all points in the restricted polytope; again, this is achieved by setting up two linear programs to maximise and minimise a particular objective, and checking whether the maximum and minimum attained are equal.

Listing 2: Output generated by Listing 1

Warning: The edge \( sy \) is actually strong up, with value \( y \)
Warning: The edge \( xy \) is actually strong up, with value \( y \)
Warning: The edge \( yz \) is actually strong up, with value \( z \)
The output (see Listing 2) to the script listed in Listing 1 indicates which tested edges the program found to be locking/strong. In particular, our assumption that \( sy \) is weak is shown to be contradicted by the set of TRI constraints we assumed to be tight; the program indicates that \( sy = y \) at all points in the restricted polytope, and so \( sy \) is actually implied to be strong, a contradiction. This means we need not consider the case where our assumed set of TRI constraints holds. As a point of interest, note that the output states that \( yz \) is forced to be equal to \( y \) and \( z \) at all points in the restricted polytope; this implies that \( yz \) is locked up, and this can indeed be verified, as demonstrated in Listing 3.

Listing 3: Demonstration of an edge which is noted to be forced into being locked up

```matlab
>> testLocking(equalities, 'yz', 'up')
Warning: y and z are locked up
> In testLocking (line 22)
```

This general approach allows us to deal efficiently with several of the checks described below. The code is available from the authors’ websites.

We indicate where this approach has been used below with the comment (verified via MATLAB program).

12.3.3 Case a)-a)

Consider the case where \( sxy \) and \( syz \) are both triangles of type 11a; see Figure 26 for an illustration. Note that \( xz \) cannot be strong down, by Lemma 24. Therefore we may take \( xz \) strong up. Note that as \( sx, sy, sz \) are all weak, we have \( \delta(x) = \delta(y) = \delta(z) = 0 \). We also note from Figure 11a that \( \delta(xy) = \delta(yz) = 0 \). Finally, note also that \( \delta(xz) = 0 \), as it strong up and its incident variables do not move. Therefore whatever TRI constraint is tight in \( xyz \), it remains tight after the perturbation, as all singleton and edge marginals do not move.

12.3.4 Case a)-b)

Consider the case where \( sxy \) is of type 11a and \( syz \) is of type 11b (so \( sz \) is the strong edge of the triangle \( syz \)); see Figure 27 for an illustration. \( xz \) can’t be strong down by Lemma 24. So we may take \( xz \) strong up. We have \( \delta(x) = 0, \delta(y) = 0, \delta(z) = -1, \) and \( \delta(xy) = 0, \delta(yz) = -1 \).

If \( z < x \), note that this implies \( \delta(xz) = \delta(z) = -1 \). There are two possible TRI inequalities that could be tight in \( xyz \). If \( x + yz = xy + xz \), then note that this TRI constraint remains tight. If \( z + sy = sx + xy \), then note in \( sxy \), either \( x + sy \geq sx + xy \) is tight - but then \( sz = x \) is strong up (verified via MATLAB program) - or \( y + sx \geq sy + xy \) is tight - but then \( sy = y \) is strong up (verified via MATLAB program), so we need not consider these cases further.

If \( z > x \), then consider triangle \( szx \): \( x < z < s \) implies \( sx \) strong up by Lemma 19 so we need not consider this case further.
For the former, we consider \( s \) or a strong edge incident to \( xz \). By symmetry of this case in \( xyz \), it suffices to consider one of these equations, and by substituting in the perturbations for each variable and edge, note that it remains tight.

For the latter, we consider \( sxy \) of type \( 11b \) (with \( sx \) strong), and \( syz \) of type \( 11b \) (with \( sz \) strong). \( xz \) can’t be strong down by Lemma 24, so we may take \( xz \) strong up. We have \( \delta(x) = -1, \delta(y) = 0, \delta(z) = -1, \) and \( \delta(xy) = -1, \delta(yz) = -1 \). Note that \( \delta(z) = -1 \) whether \( xz = x \) or \( xz = z \). The two possible tight TRI inequalities in \( xyz \) are
\[
x + yz = xy + xz \quad \text{and} \quad z + yz = xz + yz.
\]
By symmetry of this case in \( x \) and \( z \), it suffices to consider one of these equations, and by substituting in the perturbations for each variable and edge, note that it remains tight.

### 12.3.5 Case b)-b)

There are two ways in which triangles \( sxy \) and \( syz \) may be of type \( 11b \) they may share either a weak edge incident to \( s \), or a strong edge incident to \( s \).

For the former, we consider \( sxy \) of type \( 11b \) (with \( sx \) strong), and \( syz \) of type \( 11b \) (with \( sz \) strong). \( xz \) can’t be strong down by Lemma 24, so we may take \( xz \) strong up. We have \( \delta(x) = -1, \delta(y) = 0, \delta(z) = -1, \) and \( \delta(xy) = -1, \delta(yz) = -1 \). Note that \( \delta(z) = -1 \) whether \( xz = x \) or \( xz = z \). The two possible tight TRI inequalities in \( xyz \) are
\[
x + yz = xy + xz \quad \text{and} \quad z + yz = xz + yz.
\]
By symmetry of this case in \( x \) and \( z \), it suffices to consider one of these equations, and by substituting in the perturbations for each variable and edge, note that it remains tight.

For the latter, we consider \( sxy \) of type \( 11b \) and \( syz \) of type \( 11b \) (with \( sy \) the strong edge in both triangles). \( xz \) cannot be strong down by Lemma 24 so we may take \( xz \) strong up. We have \( \delta(x) = 0, \delta(y) = -1, \delta(z) = 0, \) and \( \delta(xy) = -1, \delta(yz) = -1 \). Whether \( xz = x \) or \( xz = z \), we have \( \delta(xz) = 0 \). The two possible tight TRI inequalities in \( xyz \) are
\[
x + yz = xy + xz \quad \text{and} \quad z + yz = xz + yz.
\]
By symmetry of this case in \( x \) and \( z \), it suffices to consider one of these equations, and by substituting in the perturbations for each variable and edge, note that it remains tight.

### 12.3.6 Case a)-c)

We consider \( sxy \) of type \( 11a \) \( syz \) of type \( 11c \) (so \( sz \) is the strong edge of the triangle). We have \( \delta(x) = 0, \delta(y) = 0, \delta(z) = 1, \delta(xy) = 0, \delta(yz) = 0 \). \( xz \) can’t be strong down by Lemma 24 so we may take \( xz \) strong up.

If \( xz = z \), then consider triangle \( sxz \), and note that \( s < z < x \), implying \( sx \) strong up by Lemma 19 so we don’t need to consider this case further.

If \( xz = x \), then \( \delta(xz) = \delta(x) = 0 \). There are two possible tight TRI constraints in \( xyz \). If \( x + yz = xy + xz \), then plugging in our singleton and edge perturbations immediately verifies this remains tight under the perturbation. If \( x + yz = yz + xz \), then considering triangle \( sxy \), we either have \( x + sy = xy + sx \) - in which case \( sx \) is strong up (verified via MATLAB program) - or \( y + sx = sy + xy \) - in which case \( sy \) is strong up (verified via MATLAB program), so we need not consider these cases further.

### 12.3.7 Case b)-c)

We consider \( sxy \) of type \( 11b \) (with \( sx \) strong) and \( syz \) of type \( 11c \) (with \( sz \) strong). We have \( \delta(x) = -1, \delta(y) = 0, \delta(z) = 1, \delta(xy) = -1 \) and \( \delta(yz) = 0 \). Note that \( xz \) cannot be strong down by Lemma 24 so we may take \( xz \) strong up.

If \( xz = z \), then \( s > x >= z > s \), so \( sx \) is locked up by Lemma 19 so we need not consider this case further.

If \( xz = x \), then \( \delta(xz) = \delta(x) = -1 \). There are two possible tight TRI constraints in \( xyz \). If \( z + yz = yz + xz \), then \( s \) and \( x \) are locked up (verified via MATLAB program). If \( x + yz = xy + xz \), then \( s \) and \( z \) are locked up (verified via MATLAB program).
12.3.8 Case c)-c)

There are two ways in which triangles $sxy$ and $syz$ may be of type $11c$; they may share either a weak edge incident to $s$, or a strong edge incident to $s$.

For the former, take $sxy$, $syz$ of type $11c$ with $sx$, $sz$ strong. Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up. Either $xz = x$ or $xz = z$. By symmetry of the model in $x$ and $z$, it suffices to deal with $xz = x$. Then we have $\delta(x) = 1$, $\delta(y) = 0$, $\delta(z) = 1$, and $\delta(xy) = 0$, $\delta(yz) = 0$, and $\delta(xz) = 1$. The tight TRI constraint in $xyz$ is either $x + yz = xy + xz$ or $z + xy = xz + yz$, and in both cases the perturbation keeps the TRI constraint tight.

For the latter, take $sxy$, $syz$ of type $11c$ with $sy = s$ strong. Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up. Either $xz = x$ or $xz = z$. Again by symmetry of the problem in $x$ and $z$, we need only consider $xz = x$. Then we have $\delta(x) = 0$, $\delta(y) = 1$, $\delta(z) = 0$, and $\delta(xy) = 0$, $\delta(yz) = 0$, and $\delta(xz) = 0$. There are two possible tight TRI constraints, $x + zy = xz + yz$ and $z + xy = xz + xz$ - in both cases, no terms are perturbed, so the constraints remain tight.

12.3.9 Case a)-d)

It is not possible for triangles of type $11a$ and $11d$ to share an edge incident to $s$, so we need not consider this case.

12.3.10 Case b)-d)

It is not possible for triangles of type $11b$ and $11d$ to share an edge incident to $s$, so we need not consider this case.

12.3.11 Case c)-d)

We consider $sxy$ of type $11c$ (with $sy = s$ strong up) and $syz$ of type $11d$ (with $sz = 0$ strong down). Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up.

If $xz = x$, then by considering the TRI constraint $x + sz \geq xz + sx$, we obtain $sx = 0$ strong down, a contradiction.

If $xz = z$, then note that we have $\delta(x) = 0$, $\delta(y) = 1$, $\delta(z) = 1$, and $\delta(xy) = 0$, $\delta(yz) = 0$, $\delta(xz) = 1$. There are two possible tight TRI constraints in $xyz$. If $x + yz = xy + xz$, then the above perturbation maintains the tightness of this constraint. If $z + xy = xz + yz$, this implies $sx = 0$ strong down (verified via MATLAB program), a contradiction.

12.3.12 Case d)-d)

There are two ways in which triangles $sxy$ and $syz$ may be of type $11d$, they may share either a strong up edge incident to $s$, or a strong down incident to $s$.

For the former, we consider $sxy$, $syz$ of type $11d$ (with $sy$ the strong up edge). Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up. So $xz = x$ or $xz = z$; by symmetry in $x$ and $z$, it suffices to consider $xz = x$. We have $\delta(x) = \delta(y) = \delta(z) = 1$, and $\delta(xy) = \delta(yz) = \delta(xz) = 0$, so immediately it follows that any tight TRI constraint in $xyz$ remains tight after the perturbation.

For the latter, we consider $sxy$, $syz$ of type $11d$ (with $sy$ the strong down edge). Again, we must have $xz$ strong up, and by symmetry in $x$ and $z$, it suffices to consider $xz = x$. Note that we have $\delta(x) = \delta(y) = \delta(z) = 1$, and $\delta(xy) = \delta(yz) = \delta(xz) = 0$, so immediately it follows that any tight TRI constraint in $xyz$ remains tight after the perturbation.

12.3.13 Case a)-e)

We consider $sxy$ of type $11a$ and $syz$ of type $11e$ (with $sz$ the strong down edge). Note that under a flipping of $s$, this case is the same as case a)-b).

12.3.14 Case b)-e)

We consider $sxy$ of type $11b$ (with $sz = x$ strong up), and $syz$ of type $11e$ (With $sz = 0$ strong down). By considering the TRI inequality $x + sz \geq sx + xz$, we note that $xz = 0$, and this TRI constraint is tight. The only possible tight TRI
constraint in $xyz$ is $y + xz = xy + yz$ ($x + zy = xy + xz$ implies $xy$ strong, $z + xy = xz + yz$ implies $yz$ strong, and $xy + xz + yz = x + y + z - 1$ implies $s$, $x$, $z$ are locking (verified via MATLAB program)). We have $\delta(x) = -1$, $\delta(y) = 0$, $\delta(z) = 1$, and $\delta(xy) = -1$, $\delta(yz) = 1$, $\delta(xz) = 0$. Substituting these perturbations into the tight TRI constraint $y + xz = xy + yz$, we note tightness is maintained.

12.3.15 Case c)-e)

We consider $sxy$ of type 11c (with $sx = s$ strong up) and $syz$ of type 11e (with $sz = 0$ strong down). Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up.

If $xz = x$, then the TRI inequality $x + sz \geq xs + xz$ implies $s = 0$, so we need not deal with this case.

If $xz = z$, then the only possible tight TRI constraints are $x + yz = xy + xz$ and $z + xz = xz + yz$, which both imply that $sy$ is strong (verified via MATLAB program), so we need not deal with this case.

12.3.16 Case d)-e)

We consider $sxy$ of type 11d (with $sx = s$ strong up and $sy = 0$ strong down) and $syz$ of type 11e (with $sz = 0$ strong down). Note that $xz$ cannot be strong down by Lemma 24, so take $xz$ strong up.

If $xz = x$, then $s < x < z$, so so be Lemma 19, $sx = s$ is strong up, so we need not consider this case further.

If $xz = z$, then we have $\delta(x) = 1$, $\delta(y) = 1$, $\delta(z) = 0$, and $\delta(xy) = 1$, $\delta(yz) = 1$, $\delta(xz) = 0$. There are two possible tight TRI constraints in $xyz$. If $z + xy = xz + yz$, then the perturbations described above maintain the tightness of the TRI constraint. If $x + yz = xy + xz$, this forces $sz$ to be strong up (verified via MATLAB program), so we need not consider this case further.

12.3.17 Case a)-f)

We consider $sxy$ of type 11a and $syz$ of type 11f (with $sz$ the strong down edge). Note that under a flipping of $s$, this case is the same as case a)-c).

12.3.18 Case b)-f)

We consider $sxy$ of type 11b (with $sx$ strong up), and $syz$ of type 11f (with $sz$ the strong down edge). Note that under a flipping of $s$, this case is the same as case c)-e).

12.3.19 Case c)-f)

We consider $sxy$ of type 11c (with $sx = s$ strong up), and $syz$ of type 11f (with $sz = s + z - 1$ strong down). If $xz = x$, then TRI inequality $x + sz \geq xs + sz$ implies that $sz$ is strong up, a contradiction. By Lemma 24, $xz$ cannot be strong down and equal to 0. So the cases to consider are $xz = z$ and $xz = x + z - 1$.

If $xz = z$, then note $\delta(x) = 1$, $\delta(y) = 0$, $\delta(z) = -1$, and $\delta(xy) = 0$, $\delta(yz) = 0$, $\delta(xz) = -1$. There are two possible tight TRI constraints in $xyz$. If $z + xy = xz + yz$, then the perturbation described above maintains tightness of the constraint. If $x + yz = xy + xz$, then $xyz$ is implied to be strong up (verified via MATLAB program), a contradiction.

If $xz = x + z - 1$, then the only TRI constraint that can be tight in $xyz$ is $y + xz = xy + yz$ (verified via MATLAB program). This is maintained by the perturbation described above.

12.3.20 Case d)-f)

It is not possible for triangles of type 11d and 11l to share an edge incident to $s$, so we need not consider this case.

12.3.21 Case a)-g)

It is not possible for triangles of type 11a and 11g to share an edge incident to $s$, so we need not consider this case.
12.3.22 Case b)-g)

We consider \(sx\) of type 11b (with \(sy = y\) strong up), and \(sz\) of type 11g (with \(sz = s + z - 1\) strong down). Note that \(\delta(x) = 0, \delta(y) = -1, \delta(z) = -1,\) and \(\delta(xy) = -1, \delta(yz) = -1.\) Note that \(xz\) cannot be strong down by Lemma 24, so take \(xz\) strong up.

If \(xz = x\), then \(\delta(xz) = \delta(x) = 0.\) The two possible tight TRI constraints in \(xyz\) are \(x + yz = xy + xz\) (for which it can be checked that the constraint remains tight with the perturbations specified above), and \(z + xy = xz + yz,\) which implies \(sx\) is strong down (verified via MATLAB program), a contradiction.

If \(xz = z,\) then \(\delta(xz) = \delta(z) = -1.\) The two possible tight TRI constraints in \(xyz\) are \(z + xy = xy + xz\) (for which it can be checked that the constraint remains tight with the perturbations specified above), and \(x + yz = xy + xz,\) which implies \(sx\) is strong down (verified via MATLAB program), a contradiction.

12.3.23 Case c)-g)

It is not possible for triangles of type 11c and 11g to share an edge incident to \(s,\) so we need not consider this case.

12.3.24 Case d)-g)

It is not possible for triangles of type 11d and 11g to share an edge incident to \(s,\) so we need not consider this case.

12.4 3-wedge consistency

We now consider the case where a triangle \(xyz\) not incident to \(s\) has a tight TRI constraint, and demonstrate that this TRI constraint remains tight when all singleton and edge marginals are perturbed according to the description given in §11.

We begin by arguing that we need check only 5 cases. First, the case where all edges \(sx, sy, sz\) are weak (Case 1). If two of the edges \(sx, sy, sz\) are weak, then without loss of generality we make take \(sx\) strong. Note also that if \(sx\) strong down, this is obtained from a case where \(sx\) strong up by flipping \(s,\) so we only need to consider cases where \(sx\) is strong up (Cases 2 and 3). If exactly one of the edge \(sx, sy, sz\) are weak, then without loss of generality we may take \(sx\) strong up and \(sy\) strong down. As noted in §11.2.4, there are only two cases to consider; \(sx = s, sy = 0,\) and \(sx = x, sy = s + y - 1;\) these form Cases 4 and 5. It cannot be the case that all three edges \(sx, sy, sz\) are strong, since again this would contradict Theorem 23.

12.5 Case 1

All singleton and edge marginals have 0 perturbation, so any tight TRI constraint is preserved.

12.6 Case 2

We take \(sx = s,\) all other edges weak. We note that we have \(x + sz = sx + xz\) and \(x + sy = sx + xy.\) There must also be a tight constraint in \(syz\) holding \(yz\) down, by symmetry in \(y\) and \(z\) we may take it to be \(y + sz = sy + yz.\) We then consider the four possible TRI constraints that could be tight in \(xyz.\) If \(y + xz = xy + yz,\) then the perturbation maintains the tightness of this constraint. The other three constraints lead to contradictions of tight (verified via MATLAB program).
12.7 Case 3

We take $sx = x$, all other edges weak. We note that we have $x + sz = sx + xz$ and $x + sy = sx + xy$. There must also be a tight constraint in $syz$ holding $yz$ down; by symmetry in $y$ and $z$ we may take it to be $y + sz = sy + yz$. There is also a tight constraint in $xyz$ by assumption. If it is $y + xz = xy + yz$, then the perturbation given for the singletons and edge marginals maintains the tightness of this constraint. The other three constraints lead to contradictions (verified via MATLAB program).

12.8 Case 4

We take $sx = s$, $sy = 0$, and all other edges weak. We must have $x + sz = sx + xz$, $x + sy = xy + sx$, and $z + sy = sz + yz$. We consider the four possible TRI constraints that could be tight in $xyz$. If $z + xy = xz + yz$, then the perturbation given for the singletons and edge marginals maintains the tightness of this constraint. The other three constraints lead to contradictions (verified via MATLAB program).

12.9 Case 5

We take $sx = x$, $sy = s + y - 1$, and all other edges weak. We must have the TRI constrains $x + sz = sx + xz$, $x + sy = xy + sx$, and $z + sy = sz + yz$. We consider the four possible TRI constraints that could be tight in $xyz$. If $z + xy = xz + yz$, then the prescribed perturbation works. The other three constraints lead to contradictions (verified via MATLAB program).

13 Gathering Earlier Results to Provide Proofs of Theorems 8, 9 and 11

We gather together earlier results and use them to prove the following Theorems from the main paper.

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

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

**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.

We shall first prove Theorem 11 then use it to derive Theorem 9 after which Theorem 8 will easily follow.

**Proof of Theorem 11.** uses another simple perturbation. As before, we assume an almost attractive model and as justified by §9, we assume no locking edges or variables that have singleton marginal 0 or 1. We shall prove the result by showing that, given these assumptions, the graph must have no variables other than $s$.

Given the results in §11-12, we have shown that if $s$ is fixed while other marginals are optimized, then an optimum vertex cannot occur unless the perturbation defined in §11 does not exist, i.e. we know that all other variables have a weak edge to $s$.

Hence, at an optimum vertex, there are no strong edges incident to $s$. In particular, there are no strong down edges incident to $s$, and hence there are no strong down edges anywhere in the graph (by Lemma 24).

Since there are no strong down edges, it is now easily checked that the following perturbation (times a sufficiently small $\epsilon$ s.t. all constraints which were not tight remain so) up and down preserves all tight LOC and TRI constraints:

$$
\begin{align*}
&\begin{cases}
0 & \text{if } v \in V \setminus \{s\} \\
+1 & \text{if } sv \text{ edge, with } v \in V \setminus \{s\}
\end{cases} \\
&\begin{cases}
+\frac{1}{2} & \text{if } uv \text{ edge, with } u, v \in V \setminus \{s\}
\end{cases}
\end{align*}
$$

Thus, it must be that at a vertex, all variables are either 0, 1 or in a locking component. This completes the proof of Theorem 11.

\]
Proof of Theorem 9. This is similar to the proof of Theorem 6. As there, we need only prove that $F_{\text{TRI}}^s(x)$ is convex, then linearity follows from Lemma 5.

For any $y \in [0, 1]$, consider an arg max of $F_{\text{TRI}}^i(y)$ as given by Theorem 11. Partition the variables into 4 exhaustive sets: $A_y = \{ j : q_j = 0 \}, B_y = \{ j : q_j = y \}, C_y = \{ j : q_j = 1 - y \}$ and $D_y = \{ j : q_j = 1 \}$. Define the function $f_y : [0, 1] \to \mathbb{R}$ given by $f_y(x) = f(q(x; y))$ where $q(x; y)$ is defined explicitly for singleton and edge marginals by:

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

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

It is straightforward to check that always $q(x; y) \in \text{TRI}$ (all edges are strong and there are no strong frustrated cycles). Observe that $f_y(x)$ is the linear function achieved by holding fixed the partition of variables $A_y, B_y, C_y, D_y$ that was determined for the arg max of the constrained optimum at $q_i = y$. Now $F_{\text{TRI}}^i(x) = \sup_{y \in [0, 1]} f_y(x)$, hence is convex. \[\square\]

Note that, similarly to the remark after the proof of Theorem 6, we observe that each of the $f_y(x)$ functions in the proof must be equal and hence the $A, B, C, D$ sets may be taken to be constant with the same variables in them independent of $y$.

Proof of Theorem 8. Given Theorem 9, it must be the case that a global optimum occurs at $s = 0$ or $s = 1$. If we condition on either case, the remaining model is balanced, and the result follows from Theorem 2. \[\square\]