Lifted Inference via $k$-Locality

Martin Mladenov and Kristian Kersting
Institute of Geodesy and Geoinformation, University of Bonn
Meckenheimer Allee 172, 53115 Bonn, Germany

Abstract
Lifted inference approaches exploit symmetries of a graphical model. So far, only the automorphism group of the graphical model has been proposed to formalize the symmetries used. We show that this is only the GI-complete tip of a hierarchy and that the amount of lifting depends on how local the inference algorithm is: if the LP relaxation introduces constraints involving features over at most $k$ variables, then the amount of lifting decreases monotonically with $k$. This induces a hierarchy of lifted inference algorithms, with lifted BP and MPLP at the bottom and exact inference methods at the top. In between, there are relaxations whose liftings are equitable partitions of intermediate coarseness, which all can be computed in polynomial time.

Introduction
Graphical models encountered in many AI tasks have symmetries and redundancies only implicitly captured in the graphical structure and, hence, not exploitable by efficient inference algorithms. A prominent example are relational probabilistic models that tackle a long standing goal of AI, namely unifying first-order logic (capturing regularities and symmetries) and probability (capturing uncertainty). Although they often encode large, complex models using few rules only and, hence, symmetries and redundancies abound, in the inference stage, they still operate on a mostly propositional representation level and do not exploit additional symmetries. Recently, several inference approaches that exploit symmetries have been proposed, see (Kersting 2012) for a recent overview. They essentially group together nodes indistinguishable in terms of the computations of the inference approach used, and have been proven successful in many AI tasks and applications such as information retrieval, boolean model counting, Kalman filtering, entity resolution, and biomolecular event prediction. They are often faster, more compact and provide more structure for optimization than their symmetry-agnostic counterparts.

While lifted inference approaches often yield dramatic improvements in runtime, there is still no full understanding of the underlying mathematical principles of symmetries exploited by them. Indeed, Niepert (2012) and Bui et al. (2012) have established links between the automorphism group of graphical models and lifted inference, showing that MAP-LP and MCMC can be lifted by considering the quotient of the orbit partition of the graphical model only. Unfortunately, however, computing the orbit partition is GI-complete (polynomial-time reducible to graph isomorphism in both directions). Consequently, it is an open question whether there is a polynomial-time\(^1\) algorithm for computing these “orbital” liftings. Moreover, lifted variants of approximative inference approaches such as lifted BP (Singla and Domingos 2008; Kersting, Ahmadi, and Natarajan 2009) only employ — as we will show here — the lowest level of a well-known color-refinement heuristic for graph isomorphism, called the Weisfeiler-Lehman (WL) algorithm, to compute so-called equitable partitions in (lower-order) polynomial time. These “equitable” liftings can be coarser than orbital ones but still preserve the pseudomarginals of BP. Understanding this trade-off between the degree of approximation and the computational effort required for and coarseness of lifting was the seed that grew into the present paper.

We show that automorphisms are only the GI-complete tip of a hierarchy among lifting approaches. Specifically, we show that the amount of lifting depends on how local the inference algorithm\(^2\) is: if an LP relaxation contains variables that represent a dependency among at most $k$ different vertices of a model, and the role of any variable in any constraint can be fully inferred by the connectivity among these $k$ vertices, but not their names, then the amount of lifting — the groups of variables that cannot be distinguished by the inference algorithm — refines monotonically with $k$ until it reaches the orbit partition. Intuitively, variables get grouped together if they are indistinguishable by the graph features over at most $k$ variables they are involved in. This extends recent results on local LPs (Atserias and Maneva 2013) to probabilistic inference and contributes to a deeper understanding of the interaction between symmetries and the

\(^1\)Indeed, it has been argued that for practical purposes the GI problem has been solved. The algorithms, however, do not provide any guarantees, and even polynomial algorithms for restricted classes of graphs can be prohibitively slow.

\(^2\)For the sake of simplicity, we focus on MAP inference and its linear programming relaxations. However, we note that many of the results naturally carry over to Bethe and Kikuchi free energies.
The hierarchy among lifted inference approaches is depicted in Figure 1. The bottom of our hierarchy is established by revisiting 1-MAP-LP that has been extensively studied for probabilistic inference, see e.g. (Sontag 2010). Although we will not give details here, using cycle constraints results in $|V|$-local LPs. In turn, an upper bound for indistinguishability appears to the orbit partition, which $|V|$-WL is guaranteed to find. However, using triangle constraints yields 3-local LPs, hence the orbit partition is too conservative. On the other hand, the triangle relaxation must store all equivalence classes of triplets produced by WL, in contrast to the cycle formulation that requires only pairwise marginals. Before concluding, we will touch upon an efficient lifting of $k$-local MLN inference approaches and illustrate our results.

We now exemplify the details for the intermediate levels skipping the bottom and the tip of the hierarchy due to space limitations. We refer the reader to the full paper.

The Bottom of the Lifting Hierarchy

Let us start with establishing the bottom of our hierarchy. The necessary background on MAP, (fractional) isomorphism and lifted BP is introduced on-the-fly.

**MAP Inference within Ising Models:** Let $X = (X_1, X_2, \ldots, X_n)$ be a set of $n$ discrete-valued random variables and let $x_i$ represent the possible realizations of random variable $X_i$. Markov random fields (MRFs) compactly represent a joint distribution over $X$ as a product of factors, i.e., $P(X = x) = \frac{1}{Z} \prod_i f_k(x_i)$. Here, each factor $f_k$ is a non-negative function of a subset of the variables $x_i$, and $Z$ is a normalization constant. As long as $P(X = x) > 0$ for all joint configurations $x$, the distribution can be represented as a log-linear model: $P(X = x) = \frac{1}{Z} \exp \left[ \sum_i w_i \cdot g_i(x) \right]$, where the factors $g_i(x)$ are arbitrary functions of (a subset of) the configuration $x$.

For the sake of simplicity, we will restrict our discussion to a specific subset of MRFs, namely Ising models with arbitrary topology. In an Ising model on a graph $G = (V, E)$, all variables may take only two possible states, say $\{+1, -1\}$, $i \in V$. The factors of the model specified in terms of the agreement or disagreement of variables whose nodes are adjacent, i.e. $g_{ij}(X_i = x_i, X_j = x_j) = \theta_{ij} x_i x_j$, whenever $i, j \in E$. Thus, the joint distribution is specified as $P(X = x) \propto \exp \left[ \sum_{ij \in E} \theta_{ij} \cdot x_i x_j \right]$. Now, the Maximum a-posteriori (MAP) inference problem for an Ising model is defined as finding an assignment maximizing the last equation. This can equivalently be formulated as the following linear program (LP): $\mu^* = \arg \max_{\mu \in \mathcal{M}(G)} \langle \theta, \mu \rangle$, where the set $\mathcal{M}(G)$ is the marginal polytope. Even though this is an LP, the problem of deciding membership in $\mathcal{M}(G)$ is NP-complete, hence one typically considers tractable relaxations (outer bounds) of $\mathcal{M}(G)$ and solves the resulting approximate problem.

A common outer bound on the marginal polytope is the

---

**Footnote:** The one-to-one correspondence between the parameters of the Ising model and the graph elements simplifies our arguments, sparing us some auxiliary constructions. Note however, that this restriction does not come with a loss of generality.
local polytope of doubly stochastic matrices contains nontrivial linear equations, AUT-LP constraints, this could be solved by an integer linear program. Indeed, there are broad graph families, such as perfect probabilistic inference based on 0-MAP-LP is not exact. The problem of finding an automorphism of a graph may be symmetric if it has an automorphism which is not the identity permutation. The action of all automorphisms of a graph. Aut(G) is a group under composition. The action on this group of the k-tuples of vertices and the resulting LP may admit solutions which are not valid assignments. However all integral points in M\(L(G)\) correspond to valid assignments, thus if the solution \(\mu^*\) happens to be integral, then this \(\mu^*\) is an exact solution of the MAP problem.

Capturing Symmetries: Often, we are facing inference problems with symmetries within the underlying graph structure. Specifically, a symmetry (or an automorphism) of a graph \(G = (V, E)\) is defined as a permutation \(\pi: V \rightarrow V\) such that \(ij \in E \iff \pi^T \pi E \in E\). A graph is said to be symmetric if it has an automorphism which is not the identity permutation. The set of all automorphisms of a graph, Aut(G) is a group under composition. The action of this group on the k-tuples of vertices partitions them into equivalence classes called k-orbits. We say that \((u_1, \ldots, u_r)\) is equivalent to \((v_1, \ldots, v_r)\) if there exist an automorphism \(\pi\) such that \((u_1^\pi, \ldots, u_r^\pi) = (v_1, \ldots, v_r)\). For \(k = 1\) this is a partition on the vertices which is generally referred to as the orbit partition (OP).

The problem of finding an automorphism of a graph may be stated algebraically, that is find a permutation matrix \(X\) (that is, \(X_{ij} \in \{0, 1\}\), \(X \cdot 1 = X^T \cdot 1 = 1\)) that commutes with \(A(G), XA(G) = A(G)X\), where \(A(G)\) is the (colored) adjacency matrix of \(G\). Due to the linearity of the constraints, this could be solved by an integer linear program. If we now relax the integrality constraint and require only \(X_{ij} \geq 0\), the problem becomes one of determining whether a polytope of doubly stochastic matrices contains nontrivial points (\(X \neq I_n\)). We specify the polytope by the following linear equations, AUT-LP(G) =

\[
\begin{align*}
X \geq 0 \\
\sum_{k=1}^{n} A_{ik} X_{kj} = \sum_{k=1}^{n} X_{ik} A_{kj} = 1 \\
\sum_{k=1}^{n} X_{ki} = 1
\end{align*}
\]

This relaxation is called fractional automorphism, and since it is an LP is solvable in polynomial time. Note that if we instead require \(A(G)X = XA(H)\), for a different graph \(H\), we obtain the polytope ISO-LP(G, H), a linear relaxation of the graph isomorphism problem, called fractional isomorphism, see (Ramana, Scheinerman, and Ullman 1994). Of course, it holds AUT-LP(G) = ISO-LP(G, H).

Hierarchies of Relaxations

Probabilistic inference based on 0-MAP-LP is not exact. Indeed, there are broad graph families, such as perfect graphs (Jebara 2009), where the relaxation over \(M_L(G)\) is exact. Unfortunately, there are also examples (Sontag 2010) where it fails due to having fractional vertices. In general, \(M_L(G)\) is not tight enough in many real-world problems. A common approach to tighten the approximation is to find additional constraints, which would cut away parts of the polytope but not the integer points. The hope is that after a small number of such constraints, the relevant fractional vertices would be pruned, so that an integer point can be recovered. One of these widely used tightenings for binary polytopes is due to Sherali and Adams (1990). It is applied in rounds (levels). Each round derives valid constraints on the polytope produced by the previous round, inducing a hierarchy of polytopes where every level is closer to the integer hull, where an integer solution can be found using linear programming.

The Weisfeiler-Lehman (WL) Hierarchy: It turns out that a generalized version of color-passing (CP), called the Weisfeiler-Lehman (WL) algorithm, can be used to find solutions of the different levels of the SA hierarchy of the LP relaxation of the graph isomorphism (in our case graph automorphism, AUT-LP) problem. More precisely, the k-dimensional Weisfeiler-Lehman method (k-WL) begins by partitioning (coloring) all k-tuples of vertices of a given graph \(G\) (of a graphical model). Two tuples \(u = (u_1, \ldots, u_k)\) and \(v = (v_1, \ldots, v_k)\) are assigned the same initial color, \(W^0(u) = W^0(v)\), if they have the same isomorphism type, i.e., if it holds that \((A) u_i = u_j \iff v_i = v_j\). \((B) u_i u_j \in E \iff v_i v_j \in E\), and \((C) \text{col}(u_i) = \text{col}(v_i)\).

To compute the color in iteration \(W^{r+1}(u)\), we define the operation for each \(g \in V\) and \(u \in V^k\), \(\text{sift}(f, u, g) = (f(g, u_2, \ldots, u_k), f(u_1, g, \ldots, u_k), \ldots, f(u_1, u_2, \ldots, g))\). Then, \(W^{r+1}(u) = W^{r+1}(v)\) holds if for every tuple of colors \(t\), \(|\{g \in V|\text{sift}(W^r, u, g) = t\}| = |\{g' \in V|\text{sift}(W^r, v, g') = t\}|\). The WL algorithm terminates if the coloring is stable, i.e. the partition induced by the colors does not refine anymore. See (Cai, F¨urer, and Immerman 1992) for details.

The stable partition of the k-tuples of k-WL implies a stable partition of \((k-1)\)-tuples, which is at least as fine as that of \(k-1\)-WL, since \(\text{sift}(f, (u_1, \ldots, u_k), g) = \text{sift}(f, (u_1, \ldots, u_k), g) \Rightarrow \text{sift}(f, (u_1, \ldots, u_k), g) \Rightarrow \text{sift}(f, (u_1, \ldots, u_k), g)\). Inductively, this turns into a partition of all \((< k)\)-tuples of the same length, with \(W(a) = W(b)\) if there exist \(u, v \in V^k\) with \(W(u) = W(v)\) and \(u_i = v_i, b_i = v_i\) for \(i \leq m, m\) being the length of \(a\) and \(b\). This is important since the LPs we are considering are inducing dependencies among tuples of different size.

Moreover, one actually has to consider the WL-partitions of tuples of vertices of colored hypergraphs. For the sake of simplicity, however, we stay with a simple graph representation. This is no restriction, since the definition of isomorphism type may be easily extended to reflect this additional structure. Moreover, keep in mind that a colored, oriented

\[\text{Note that Atserias and Maneva ((Atserias and Maneva 2013)) partition the (< k)-tuples by introducing a placeholder vertex \(*\) and requiring that \(a_i = u_i, b_i = v_i\) for \(i \leq m\) and \(v_i = u_i = *\) for \(m < i \leq k\). These two approaches are equivalent.}\]
Another notational convenience is that we will allow multi-vertex hyperedges. Thus, all our constructions can be cast in terms of ordinary graphs. Another notational convenience is that we will allow multiple colors per node, i.e. the \( \text{col}(\cdot) \) function now returns a set.

Transfer between SA and WL: Indeed, \( k \)-WL can be used to compute the OP for many graph families by projecting \( k \)-WL (for some small \( k \)) onto \( V \). In turn, one could use \( k \)-WL for realizing the OP-based lifted inference approaches in (Bui, Huynh, and Riedel 2012; Niepert 2012). There are, however, hard cases (Cai, Fürer, and Immerman 1992) for which \( O(|V|) \)-WL is needed, making it an exponential algorithm for computing the OP in general. And, we are actually not interested in the OP but rather in the sequence of partitions on the \( k \)-tuples of a graph \( G \) produced by \( k \)-WL and how they induce a hierarchy among lifted inference approaches. For this, we employ that the levels of the SA hierarchy applied to fractional automorphism interchange in power with the levels of the WL hierarchy.

Recall that the result of projecting a partition of \( V^k \) to \( V \) is equitable on \( V \), and hence can be used as a solution to AUT-LP. For \( k = 1 \), this is the coarsest equitable partition (Ramana, Scheinerman, and Ullman 1994). A recent striking result due to Atserias and Maneva (2013) shows that the \( k \)-WL partitions of \( V^k \) have a polyhedral interpretation, too. That is, the \( k \)-WL partition on \( V^k \) can be turned into a feasible point of the \( k \)-level polytope of the SA hierarchy of AUT-LP. Originally, Atserias and Maneva worked out the explicit form for the \( k \)-ISO-LP. Here, we will adopt it for \( k \)-AUT-LP:

\[
\begin{align*}
X \geq 0 & \quad \sum_{k=1}^{n} A_{ik} X_{jj(k,j)} = \sum_{k=1}^{n} X_{jj(i,k)} A_{kj} \\
\sum_{k=1}^{n} X_{ij(i,k)} &= X_q \\
X_{0} &= 1
\end{align*}
\]

where \( q \) is a set of at most \( k \) pairs of vertices of \( G \), i.e. \( q \subseteq V^2 \), \( |q| \leq k \). For our purposes, we can consider the subsets of \( V^2 \) to represent partial mappings between two \( r \)-tuples of vertices of \( A \). That is, if \( p = \{(u_1, v_1), \ldots, (u_r, v_r)\} \), then \( p(u_i) = v_i \). Note that not all \( p \)'s in the formulation of \( k \)-AUT-LP represent partial mappings. In the solution we are interested in, however, if \( p \) is not a partial mapping, then \( X_p = 0 \). Adopting the notation of (Atserias and Maneva 2013), we will represent a partial map \( p \) between \( u \) and \( v \) as \( u \mapsto v \). Now, applying Atserias and Maneva’s Transfer Lemma between SA and WL yields the following solution to \( k \)-AUT-LP(G):

\[
X_p = \begin{cases} 
\frac{1}{W(v)} & \text{if } p \equiv v \mapsto u \text{ and } W(v) = W(u) \\
0 & \text{otherwise.}
\end{cases}
\]  

Note that since \( q \) is an unordered set of pairs, if \( p \) represents \( u \mapsto v \), then it also represents \( \pi \circ u \mapsto \pi \circ v \) for all permutations \( \pi \). However, the assignment of \( X_p \) is still well-defined. Since \( \text{sift}(f, u, g) = \text{sift}(f, v, g') \Rightarrow \text{sift}(f, \pi \circ u, g) = \text{sift}(f, \pi \circ v, g') \), we may be certain that \( W(u) = W(v) \Rightarrow W(\pi \circ u) = W(\pi \circ v) \), even though \( W(\pi) \) may not be equal to \( W(\pi) \). The reason why this result is important for us, is that, as Atserias and Maneva show, the solutions of \( k \)-local linear programs are preserved by the solutions of \( k \)-AUT-LP. In the following we review what \( k \)-locality means.

Let \( G \) be a graph. We say that the size \( |u| \) of the \( k \)-tuple of vertices \( u \) is the number of distinct vertices contained in \( u \). We define the map \( \gamma_u : \{u_1, \ldots, u_k\} \rightarrow \{1, \ldots, |u|\} \) to be the unique bijection such that \( \gamma_u(u_i) \leq \sum_j \gamma_u(u_j) \). In other words, \( \gamma_u \) arranges the unique elements of \( u \) in order of their first appearance. By \( [G, u] \) we denote the pair of the “generic” colored graph, which is isomorphic to the subgraph of \( G \) induced by elements of \( u \), together with its order-tuple. More specifically, \([G, u] \) contains: (i) \( \{u_1, \ldots, u_k\} \), (ii) \( \{\gamma_u(u_1), \ldots, \gamma_u(u_k)\} : (u_i, u_j) \in G \}, \) (iii) colors \( \text{co}(G,u)(\gamma_u(u_i)) = \text{col}(u_i) \), and (iv) order-tuple \( \{(\gamma_u(u_1), \ldots, \gamma_u(u_k)) \mid \text{order-tuple} \} \). An LP \( \mathcal{L}(G) \) derived from a graph \( G \) having one variable \( x_u \) for every \((\leq k)\)-tuple of vertices \( u \) and one constraint for every \((\leq k)\)-tuple \( v \), is called a basic \( k \)-local LP, if it can be written as:

\[
\sum_{r=1}^{k} \sum_{u,v \in V, \text{order-tuple}} M_{ruv} x_u \leq d_{r,v}.
\]

That is, every constraint must be constructed using only the information found in the generic induced subgraph of \( G \) having at most \( k \) vertices, as illustrated in Fig. 2. Now, an LP is \( k \)-local if it is the union of basic \( k \)-local LPs. Finally, to prove our results, we need the following theorem from (Atserias and Maneva 2013), which essentially says that \( k \)-local LPs cannot distinguish \( k \)-SA isomorphic:

**Theorem 1** Let \( G = (V, E) \) and \( H = (U, F) \) be graphs such that \( G \equiv_{SA} H \) and let \( X \) be a solution to \( k \)-ISO-LP(G,H) witnessing that fact. Then, \( \mathcal{L}(G) \) is feasible iff
Algorithm 1: Lifted Clustered MAP LP

1. Construct $k$-MAP-LP($G$) = $(A, b, c)$ of an Ising model $G$;
2. Determine its indistinguishable tuples using $(k + 2)$-WL($G$);
3. Read off the block matrix $B$;
4. Obtain the solution $r$ of the LP $(AB, b, BTc)$ using any standard LP solver;
5. Return MAP beliefs $x^* = Br$;

$\mathcal{L}(H)$ is feasible. If the vector $x^* = (x_v)_{v \in V, r \leq k}$ is a solution to $\mathcal{L}(G)$, then $y^* = (y_u)_{u \in U,r \leq k}$ with $y_u = \sum_{v \in V} x_v$ is a solution to $\mathcal{L}(H)$.

Since this also holds for fractional automorphisms, the distinguishing power at level $k < |V|$ induces a hierarchy among lifted inference approaches, with moving up the hierarchy generally resulting in better approximations.

Lifting Hierarchy

More formally, for $k$-AUT-LP($G$) = $k$-ISO-LP($G, G$), Theorem 1 tells us that the subspace spanned by the columns of $X$ contains a solution. Recall now that $X$ can be constructed out of the $k$-WL partition by means of Eq. 4. In fact, let us think about the linear function $x_u \mapsto \sum_{v \in V} x_v$ in terms of matrix/vector multiplication in the vector space $\mathbb{R}^V$, with $V = \bigcup_{r=0}^k V^r$. The matrix corresponding to this operation has rows and columns indexed by the $(\leq k)$-tuples of $V$, i.e. $X = (X_{uv})_{u \in U, v \in V}$. Moreover, by the construction of Eq. 4, $X_{uv} = X_{u \rightarrow v} = \frac{1}{|W(u)|}$ if $W(u) = W(v)$ and 0 otherwise. Note that condition (i) on $W^0$ implies that the equivalence classes of $WL$ consist only of those tuples among which a partial mapping is possible. One can now verify that $X = BB^T$ for

$$E_{un} = \begin{cases} \frac{1}{|P_u|^2} & \text{if tuple } u \text{ belongs to some part } P_u, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, similarly to the 1-dimensional case, the bottom of our hierarchy, we solve $\mathcal{L}(G)$ over the space defined by the equivalence classes, i.e., supervariables resulting from running $k$-WL. This establishes higher levels of a hierarchy among lifted inference algorithms as summarized in Alg. 1 and proves the following theorem:

**Theorem 2** For any $k$-local LP, a partition of indistinguishable variables is computed by $k$-WL. Furthermore, the coarseness of the partitions decreases monotonically with $k$ until it reaches the OP.

To illustrate this, we show that MAP over $M_{L}^{(0,1)}(G)$ is 2-local. Let $v = (i,j)$; for constraints of the type $\mu_{ij} \leq \mu_i$, we have $M_{L}^{G,uv} = 1$ if the order tuple of $uv$ is $(1,2,1,2)$ (meaning $u = v$) and there is an edge between $\gamma_v(i)$ and $\gamma_v(j)$ in the induced generic subgraph of $uv$ in $G$. Correspondingly, $M_{L}^{G,uv} = -1$ if the order tuple of $uv$ is $(1,1,2)$ (meaning $u = (i)$) and $(1,2)$ is an edge in $[G, uv]$.

We can similarly define the constraint $\mu_i + \mu_j - \mu_{ij} \leq 1$ as $M_{L}^{G,uv} = 1$ for order tuples $(1,1,2)$, corresponding to $uv = (i,i,j)$ and $(1,2,1)$, corresponding to $uv = (j,i,j)$; $M_{L}^{G,uv} = -1$ for order tuple $(1,2,1)$ and $d_{G,v} = 1$ if $\gamma_i(v), \gamma_j(v) \in [G, v]$. Finally, the objective, which can be expressed as a constraint $\sum_{i \in E} \theta_i \mu_{ij} \leq W$ has a local representation as $M_{L}^{G,uv} = \theta_{col}(uv)$ (parameters are encoded as edge colors) if $v$ is the empty tuple and the generic graph of $u$ has an edge, $d_{G,v} = 1$ is then $-W$.

In general, Atserias and Maneva have shown that the classical LP relaxations of combinatorial problems such as bipartite matchings and maximum flows are 2-local. Furthermore, they examined the SA-hierarchy of the max-cut relaxation over the metric polytope and found the $k$-th level to be $(2k + 1)$-local. It immediately follows that we have a $(2k + 1)$-local tightening of $M_{L}^{(0,1)}(G)$, since $M_{L}^{(0,1)}(G)$ is affinely equivalent to the Metric Polytope restricted to particular edges (also called rooted metric polytope) (Sontag 2010). However, please note that there is a slight difference between what we obtain by applying SA as defined above to $M_{L}^{(0,1)}(G)$ and what is commonly referred to as the SA-hierarchy of the local polytope in the graphical models literature, see e.g. (Sontag 2010). Namely, the k-SA of $M_{L}(G)$ is typically obtained by adding constraints of the type $\sum_{x_{ij} \in \Gamma} \gamma_i(x_c) = \mu_{ij}(x_i, x_j)$ for all $c \subset V, c \subseteq \{i, j\}, |c| \leq k + 2$. The difference lies in the fact that the size of the constraining set of variables increases by 1 for every subsequent $k$; whereas by applying our definition of SA, we have to increase the index set by 2 as we are multiplying constraints with edge marginals. The former would be obtained only adding the constraints $\mu_i(a^T \mu - b)$ and not $\mu_{ij}(a^T \mu - b)$. An induction argument shows that the tightening of $M_{L}^{(0,1)}(G)$ obtained in this way is $(k + 2)$-local.

**Theorem 3** A lifted network, i.e., a partition of indistinguishable variables of $k$-SA-MAP($G$) can be found in polynomial time for a fixed $k$. Moreover, this partition is at least as coarse as the OP of $G$.

This can be seen as follows. From our discussion above, it follows that the equivalence classes of $k$-WL yield indistinguishable variables in $k$-SA-MAP($G$). The running time of $k$-WL is $O(k^2 n^{k+1} \log(n))$, see e.g. (Cai, Fürer, and Immerman 1992), which agrees with the running time of colorpassing (see above) since the latter one coincides with 1-WL. Since the $k$-WL method does not solve isomorphism for a constant $k$, it will also yield partitions which are always as coarse as the $k$-orbits of the graph, often coarser.

Illustration

To illustrate Alg. 1, we used the so-called Frucht (among 12 people) and McKay (among 8 people) graphs, see Fig. 3, to encode the social network in a binary version of the Smokers MLN. They induced ground 0-MAP-LPs with 456 variables and 1848 constraints resp. 208 variables and 826 constraints. Fig. 4(a) summarizes the sizes of the corresponding lifted LPs. The size of the lifted Frucht graph LP using 0-WL is significantly smaller than the ground LP, which for
Conclusions

We have established a hierarchy of lifted inference approaches. It explores the space between lifted BP and lifted exact inference. The central underlying ideas are that of local linear programs, a concept so far not used to characterize probabilistic inference, and a recent deep link between the Weisfeiler-Lehman and the Sherali-Adams (SA) hierarchies provided for local LPs. The hierarchy shows that tractable lifting approaches exist for any finite LP tightening via SA and even for exact inference problems with small treewidth. Since, our results do not depend on a relational specification of the model, more broadly, we establish a connection between locality and probabilistic inference in general. Exploring this connection further, e.g., for characterizing the complexity of probabilistic inference in terms of locality is the most attractive avenue for future work. Our results also suggest an affirmative answer to the open question whether all MAP message-passing algorithms are liftable: by solving the lifted LPs using coordinate descent in the dual, one could turn the hierarchy of LP relaxations into lifted message-passing algorithms.

Acknowledgments

This work was supported by the German Research Foundation DFG, KE 1686/2-1, within the SPP 1527, and the German-Israeli Foundation for Scientific Research and Development, 1180-218.6/2011.

References