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 pseudo-marginals 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.

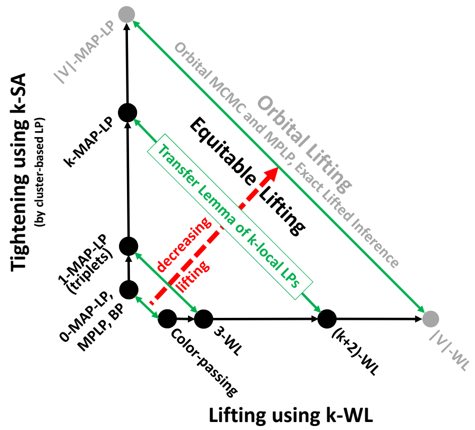

Figure 1: The hierarchy among lifted inference approaches induced by k-locality of LPs. At the lowest level, there are lifted variants of BP and MPLP, which constrain only a single variable and one of its neighbors at a time. At the top, there are e.g. cycle relaxation of MAP, where any number of variables may participate in a cut constraint, and exact inference approaches. Thus, the variables could potentially be distinguished only up to the symmetry group of the graphical model, forcing one to revert to orbit partitions. In between, there are relaxations whose distinguishing powers are characterized by various equitable partitions of intermediate coarseness induced by k-locality. In a sense, the length for the green lines denotes the distinguishing power of the algorithm. (Best viewed in color)

complexity of inference: (1) It establishes the first hierarchy of lifted inference algorithms as shown in Fig. 1. (2) For k-local inference approaches, a fixed number of WL levels suffice to determine a lifted network. The network is computable in polynomial time. (3) Since the WL hierarchy and the Sherali-Adams (SA) hierachy of LP relaxations interleave, the tightening of MAP relaxations via a fixed number of SA levels is preserved by lifting.

What follows is motivated by the equivalence of color-passing for lifting BP and the 1-dimensional WL level. The equivalence is due to a result by Ramana *et al.* (1994) that color-passing solves an LP relaxation of isomorphism (ISO-LP) called fractional isomorphism.As fractional automorphisms (AUT-LP) of graphical models preserve the MAP solutions over the local polytope, this characterizes the lifting used in lifted BP, since the feasible set of BP is the local polytope itself. Motivated by this, we then **build up the hierarchy** by using higher levels of WL (k-WL). Specifically, we investigate SA-based tightenings of LP relaxations (Sontag, Globerson, and Jaakkola 2008), which we call k-MAP-LP, and employ deep results due to Atserias and Maneva (2013) on local LPs and on k-WL interleaving with the k-th level of SA (which we call k-AUT-LP) to prove that k-WL yields a valid lifting for SA tightenings of MAP. Specifically, we prove that the family of LPs at the kth level of SA of MAP-LP, i.e., k-MAP-LP is $(k+2)$-local. Thus, the $k+2$ WL levels yields the partition of indistinguishable variables.

Finally, one can establish the **tip of our hierarchy** 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 **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 $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ 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 $\mathbf{X}$ as a product of factors, i.e., $P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_k f_k(\mathbf{x}_k)$. Here, each factor f_k is a non-negative function of a subset of the variables $\mathbf{x}_k$, and Z is a normalization constant. As long as $P(\mathbf{X} = \mathbf{x}) > 0$ for all joint configurations $\mathbf{x}$, the distribution can be represented as a log-linear model: $P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left[\sum_i w_i \cdot g_i(\mathbf{x})\right]$, where the factors $g_i(\mathbf{x})$ are arbitrary functions of (a subset of) the configuration $\mathbf{x}$.

For the sake of simplicity, we will restrict our discussion to a specific subset of MRFs, namely Ising models with arbitrary topology[3]. In an Ising model on a graph $G = (V, E)$, all variables may take only two possible states, say $\mathbf{X}_i \in \{-1, 1\}, i \in V$. The factors of the model specified in terms of the agreement or disagreement of variables whose nodes are adjacent, i.e. $g_{ij}(\mathbf{X_i} = x_i, \mathbf{X_j} = x_j) = \theta_{ij} x_i x_j$, whenever $i, j \in E$. Thus, the joint distribution is specified as $P(\mathbf{X} = \mathbf{x}) \propto \exp\left[\sum_{ij \in E} \theta_{ij} \cdot x_i x_j\right]$. 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) $\boldsymbol{\mu}^* = \operatorname{argmax}_{\boldsymbol{\mu} \in \mathcal{M}(G)} \langle \boldsymbol{\theta}, \boldsymbol{\mu} \rangle$, 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

[3]The one-to-one correspondence between the parameters of the Ising model and the graph elements simplifies our arguments to, sparing us some auxiliary constructions. Note however, that this restriction does not come with a loss of generality.

local polytope, see e.g. (Sontag and Jaakkola 2007; Sontag 2010) and references in there, defined for a pairwise model as follows, $\mathcal{M}_L(G) =$

$$\left\{ \boldsymbol{\mu} \geq \mathbf{0} \;\middle|\; \begin{array}{l} \sum_{s \in \mathbf{X}_i} \mu_{ij;st} = \mu_{j;t}, ij \in E \\ \sum_{s \in \mathbf{X}_i} \mu_{i;s} = 1, i \in V \end{array} \right\}. \qquad (1)$$

For Ising models, Eq. 1 can be projected onto the marginals $\mu_{ij;11}$ and $\mu_{i;1}$ resulting in the formulation, which we call the binary local polytope, see also (Sontag 2010),

$$\mathcal{M}_L^{\{0,1\}}(G) = \left\{ \boldsymbol{\mu} \geq \mathbf{0} \;\middle|\; \begin{array}{l} \forall ij \in E : \\ \mu_{ij} \leq \mu_j, \mu_{ij} \leq \mu_i \\ \mu_i + \mu_j - \mu_{ij} \leq 2 \end{array} \right\}. \qquad (2)$$

Note that whenever $\mathcal{M}(G)$ and $\mathcal{M}_L(G)$ do not coincide, $\mathcal{M}_L(G)$ (which is a superset of $\mathcal{M}(G)$) has fractional vertices and the resulting LP may admit solutions which are not valid assignments. However all integral points in $\mathcal{M}_L(G)$ correspond to valid assignments, thus if the solution $\boldsymbol{\mu}^*$ happens to be integral, then this $\boldsymbol{\mu}^*$ 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 $\pi : V \to V$ such that $ij \in E \Leftrightarrow u^\pi v^\pi \in E$. 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 of a graph partitions them into equivalence classes called k-orbits. We say that $(u_1, \ldots, u_r)$ is equivalent to $(v_1, ..., v_r)$ iff there exist an automorphism π such that $(u_1^\pi, \ldots, u_k^\pi) = (v_1, \ldots, v_r)$. 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 $\mathbf{A}(G), \mathbf{X}\mathbf{A}(G) = \mathbf{A}(G)\mathbf{X}$, where $\mathbf{A}(G)$ 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$(G) =$

$$\left\{ \mathbf{X} \geq \mathbf{0} \;\middle|\; \begin{array}{l} \sum_{k=1}^n A_{ik} X_{kj} = \sum_{k=1}^n X_{ik} A_{kj} \\ \sum_{k=1}^n X_{ik} = \sum_{k=1}^n X_{ki} = 1 \end{array} \right\}. \qquad (3)$$

This relaxation is called fractional automorphism, and since it is an LP is solvable in polynomial time. Note that if we instead require $\mathbf{A}(G)\mathbf{X} = \mathbf{X}\mathbf{A}(H)$, 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, G) .

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 $\mathcal{M}_L(G)$ is exact. Unfortunately, there are also examples (Sontag 2010) where it fails due to having fractional vertices. In general, $\mathcal{M}_L(G)$ 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 $\mathbf{u} = (u_1, \ldots, u_k)$ and $\mathbf{v} = (v_1, \ldots, v_k)$ are assigned the same initial color, $W^0(\mathbf{u}) = W^0(\mathbf{v})$, if they have the same isomorphism type, i.e., iff it holds that **(A)** $u_i = u_j \Leftrightarrow v_i = v_j$, **(B)** $u_i u_j \in E \Leftrightarrow v_i v_j \in E$, and **(C)** $\mathrm{col}(u_i) = \mathrm{col}(v_i)$. To compute the color in iteration $W^{r+1}(\mathbf{u})$, we define the operation for each $g \in V$ and $\mathbf{u} \in V^k$: $\mathrm{sift}(f, \mathbf{u}, g) = (f(g, u_2, \ldots, u_k), f(u_1, g, \ldots, u_k), \ldots, f(u_1, u_2, \ldots, g))$. Then, $W^{r+1}(\mathbf{u}) = W^{r+1}(\mathbf{v})$ holds if for every tuple of colors $\mathbf{t}$, $|\{g \in V | \, \mathrm{sift}(W^r, \mathbf{u}, g) = \mathbf{t}\}| = |\{g' \in V | \, \mathrm{sift}(W^r, \mathbf{v}, g') = \mathbf{t}\}|$. The WL algorithm terminates if the coloring is stable, i.e. the partition induced by the colors does not refine anymore. See (Cai, Fürer, 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 $\mathrm{sift}(f, (u_1, \ldots, u_k), g) = \mathrm{sift}(f, (v_1, \ldots, v_k), g') \Rightarrow \mathrm{sift}(f, (u_1, \ldots, u_{k-1}), g) = \mathrm{sift}(f, (v_1, \ldots, v_{k-1}), g')$. Inductively, this turns into a partition[4] of all $(< k)$-tuples of the same length, with $W(\mathbf{a}) = W(\mathbf{b})$ if there exist $\mathbf{u}, \mathbf{v} \in V^k$ with $W(\mathbf{u}) = W(\mathbf{v})$ and $a_i = u_i, b_i = v_i$ for $i \leq m$, m being the length of $\mathbf{a}$ and $\mathbf{b}$. 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

[4]Note that Atserias and Maneva ((Atserias and Maneva 2013)) partition the $(< k)$-tuples by introducing a placeholder vertex $\star$ and requiring that $a_i = u_i, b_i = v_i$ for $i \leq m$ and $v_i = u_i = \star$ for $m < i \leq k$. These two approaches are equivalent.

hypergraph can be uniquely converted to a colored simple graph preserving all topological information by the addition of linearly many extra vertices. Thus, all our constructions that use hypergraphs can be cast in terms of ordinary graphs. Another notational convenience is that we will allow multiple colors per node, i.e. the $\mathrm{col}(\cdot)$ 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 interleave 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[5]: k-AUT-LP(G) =

$$\left\{ \mathbf{X} \geq 0 \;\middle|\; \begin{array}{l} \sum_{k=1}^{n} A_{ik} X_{q\cup(k,j)} = \sum_{k=1}^{n} X_{q\cup(i,k)} A_{kj} \\ \sum_{k=1}^{n} X_{q\cup(i,k)} = X_q \\ \sum_{k=1}^{n} X_{q\cup(k,j)} = X_q \\ X_{\emptyset} = 1 \end{array} \right\}.$$

Here, 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 $\mathbf{u}$ and $\mathbf{v}$ as $p = \mathbf{u} \mapsto \mathbf{v}$. Now, applying Asterias and Maneva's Transfer Lemma between SA and WL yields the following solution to k-AUT-LP(G):

$$X_p = \begin{cases} \frac{1}{|W(\mathbf{v})|} & \text{if } p = \mathbf{v} \mapsto \mathbf{u} \text{ and} \\ & W(\mathbf{v}) = W(\mathbf{u}), \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Note that since q is an unordered set of pairs, if p represents $v \mapsto u$, then it also represents $\pi \circ v \mapsto \pi \circ u$ for all permutations π. However, the assignment of X_p is

[5] Due to space restrictions, we are omitting some details and refer the reader to (Atserias and Maneva 2013).

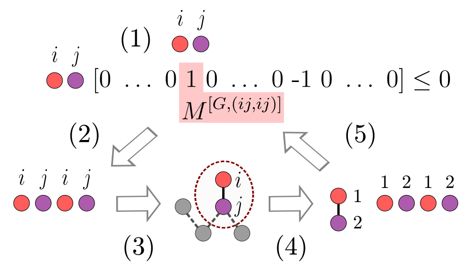

Figure 2: Illustration of locality: (1) Each entry in the constraint matrix is indexed by tuples of nodes. (2) For each entry independently, its row and column indices are concatenated. (3) Then, we consider the subgraph induced by the unique sets of nodes appearing in the concatenation and (4) anonymize it and the concatenation. (5) If we can decide on the entry (in our case that it is 1) only based on the anonymized information, this constraint is basic and local. Intuitively, variables get grouped together if they are indistinguishable by the graph features over at most k variables they are involved in. (Best viewed in color)

still well-defined. Since $\mathrm{sift}(f, \mathbf{u}, g) = \mathrm{sift}(f, \mathbf{v}, g') \Rightarrow \mathrm{sift}(f, \pi \circ \mathbf{u}, g) = \mathrm{sift}(f, \pi \circ \mathbf{v}, g')$, we may be certain that $W(\mathbf{u}) = W(\mathbf{v}) \Rightarrow W(\pi \circ \mathbf{u}) = W(\pi \circ \mathbf{v})$, even though $W(\mathbf{u})$ may not be equal to $W(\pi \circ \mathbf{u})$. The reason why this result is important for us, is that, as Asterias 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 $\mathbf{u}$ is the number of distinct vertices contained in it. We define the map $\gamma_{\mathbf{u}} : \{u_1, \ldots, u_k\} \rightarrow \{1, \ldots, |u|\}$ to be the unique bijection such that $\gamma_{\mathbf{u}}(u_i) \leq |(u_1, \ldots, u_i)|$. In other words, $\gamma_{\mathbf{u}}$ arranges the unique elements of $\mathbf{u}$ by order of their first appearance. By $[G, \mathbf{g}]$ we denote the pair of the "generic" colored graph, which is isomorphic to the subgraph of G induced by elements of $\mathbf{u}$, together with its order-tuple. More specifically, $[G, \mathbf{u}]$ contains: (i) vertices $\{1, ..., |\mathbf{u}|\}$, (ii) edges $\{(\gamma_{\mathbf{u}}(u_i), \gamma_{\mathbf{u}}(u_j)) : (u_i, u_j) \in G\}$, (iii) colors $\mathrm{col}^{[G,\mathbf{u}]}(\gamma_{\mathbf{u}}(u_i)) = \mathrm{col}(u_i)$, and (iv) order-tuple $(\gamma_{\mathbf{u}}(u_1), \gamma_{\mathbf{u}}(u_2), ..., \gamma_{\mathbf{u}}(u_k))$. 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:

$$\sum_{r=1}^{k} \sum_{\substack{\mathbf{u} \in V^r, \\ |\mathbf{uv}| \leq k}} M_r^{[G,\mathbf{uv}]} x_{\mathbf{u}} \leq d^{[G,\mathbf{v}]} \,. \tag{5}$$

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 $G \equiv_{SA}^{k} H$ 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)= $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ of an Ising model G;
2. Determine its indistinguishable tuples using $(k+2)$-WL(G);
3. Read off the block matrix $\mathcal{B}$;
4. Obtain the solution $\mathbf{r}$ of the LP $(\mathbf{A}\mathcal{B}, \mathbf{b}, \mathcal{B}^T\mathbf{c})$ using any standard LP solver;
5. **return** *MAP beliefs* $\mathbf{x}^* = \mathcal{B}\mathbf{r}$;

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

Since this also holds for fractional automorphisms, the distinguishing power at level $k < |V|$ induce 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 $x_\mathbf{u} \mapsto \sum_{\mathbf{v}\in V^r} X_{\mathbf{v}\mapsto\mathbf{u}} x_\mathbf{v}$ in terms of matrix/vector multiplication in the vector space $\mathbb{R}^\mathcal{V}$, with $\mathcal{V} = \bigcup_{r=0}^k V^r$. The matrix corresponding to this operation has rows and columns indexed by the $(\leq k)$-tuples of V, i.e. $\mathcal{X} = (X_{\mathbf{uv}})_{\mathbf{u},\mathbf{v}\in\mathcal{V}}$. Moreover, by the construction of $Eq.$ 4, $X_{\mathbf{uv}} = X_{\mathbf{u}\mapsto\mathbf{v}} = \frac{1}{|W(\mathbf{u})|}$ if $W(\mathbf{u}) = W(\mathbf{v})$ 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 $\mathcal{X} = \mathcal{B}\mathcal{B}^T$ for

$$\mathcal{B}_{\mathbf{u}n} = \begin{cases} \frac{1}{|P_n|^{\frac{1}{2}}} & \text{if tuple } \mathbf{u} \text{ belongs to some part } P_n \\ 0 & \text{otherwise.} \end{cases}$$

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 $\mathcal{M}_L^{\{0,1\}}(G)$ is 2-local. Let $\mathbf{v} = (i, j)$; for constraints of the type $\mu_{ij} \leq \mu_i$, we have $M^{[G,\mathbf{uv}]} = 1$ if the order tuple of $\mathbf{uv}$ is $(1, 2, 1, 2)$ (meaning $u = v$) and there is an edge between $\gamma_\mathbf{uv}(i)$ and $\gamma_\mathbf{uv}(j)$ in the induced generic subgraph of $\mathbf{uv}$ in G. Correspondingly, $M^{[G,\mathbf{uv}]} = -1$ if the order tuple of $\mathbf{uv}$ is $(1, 1, 2)$ (meaning $\mathbf{u} = (i)$) and $(1, 2)$ is an edge in $[G, \mathbf{uv}]$. We can similarly define the constraint $\mu_i + \mu_j - \mu_{ij} \leq 1$ as $M^{[G,\mathbf{uv}]} = 1$ for order tuples $(1, 1, 2)$, corresponding to $\mathbf{uv} = (i, i, j)$ and $(1, 2, 1)$, corresponding to $\mathbf{uv} = (j, i, j)$, $M^{[G,\mathbf{uv}]} = -1$ for order tuple $(1, 2, 1, 2)$ and $d^{[G,\mathbf{v}]} = 1$ if $\{\gamma_\mathbf{v}(i), \gamma_\mathbf{v}(j)\} \in [G, \mathbf{v}]$. Finally, the objective, which can be expressed as a constraint $\sum_{ij\in E} \theta_{ij}\mu_{ij} \leq W$ has a local representation as $M^{[G,\mathbf{uv}]} = \theta_{col(\mathbf{uv})}$ (parameters are encoded as edge colors) if $\mathbf{v}$ is the empty tuple and the generic graph of $\mathbf{u}$ has an edge, $d^{[G,\mathbf{v}]}$ 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 $\mathcal{M}_L^{\{0,1\}}(G)$, since $\mathcal{M}_L^{\{0,1\}}(G)$ 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 $\mathcal{M}_L^{\{0,1\}}(G)$ 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 $\mathcal{M}_L(G)$ is typically obtained by adding constraints of the type $\sum_{x_{c\setminus i,j}} \tau_c(x_c) = \mu_{ij}(x_i, x_j)$ for all $c \subseteq V, c \supseteq \{i, j\}, |c| \leq k+2$. 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 $\mu_i(\mathbf{a}^T\boldsymbol{\mu} - b)$ and not $\mu_{ij}(\mathbf{a}^T\boldsymbol{\mu} - b)$. An induction argument shows that the tightening of $\mathcal{M}_L^{\{0,1\}}(G)$ 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 $\mathcal{O}(k^2 n^{k+1} \log(n))$, see e.g. (Cai, Fürer, and Immerman 1992), which agrees with the running time of color-passing (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 than 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 variable 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

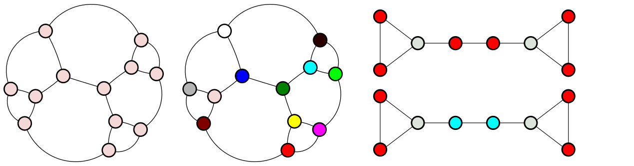

Figure 3: (Left) The Frucht graph with 12 nodes, and (right) the McKay graph with 8 nodes. The colors indicate the resulting node partitions (of these graphs but not for the corresponding MAP-LPs) for 1-WL and 2-WL, which coincide with the EP resp. OP. As one can see, 1-WL cannot distinguish any nodes in the Frucht graph, but 2-WL distinguishes all. For McKay, even 2-WL does not distinguish all the nodes due to symmetries. Thus, moving up the hierarchy, i.e., getting more and more exact solutions may produce smaller liftings. (Best viewed in color)

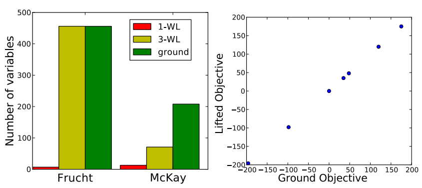

Figure 4: Empirical illustration. From left to right: (a) Sizes of ground and lifted LPs. (b) Achieved objectives. (Best viewed in color)

the Frucht graph coincides with the OP [6]. We also moved up the hierarchy, computing 1-MAP-LP. This induced ground LPs with up to 85400 variables and 670036 contraints. We computed 1-MAP-LP using 3-WL on the grounded LP. Finally, for all experiments so far we recorded the achieved objective values for the ground and lifted LPs. Additionally, we computed 1-MAP-LP (ground and lifted) for the Smokers MLN with no evidence for different clause weights. The performances are summarized in Fig. 4(b); the achieved objectives always coincided.

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.

[6] Generally, there are graphs, where we need to run $|V|$-WL to produce the OP (Cai, Fürer, and Immerman 1992).

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

Atserias, A., and Maneva, E. 2013. Sherali-Adams relaxations and indistinguishability in counting logics. *SIAM Journal of Computation* 42(1):112–137.

Bui, H.; Huynh, T.; and Riedel, S. 2012. Automorphism groups of graphical models and lifted variational inference. *CoRR* abs/1207.4814.

Cai, J.-Y.; Fürer, M.; and Immerman, N. 1992. An optimal lower bound on the number of variables for graph identification. *Combinatorica* 12:389–410.

Jebara, T. 2009. MAP estimation, message passing, and perfect graphs. In *Proc. of the 25th Conference on Uncertainty in Artificial Intelligence (UAI-2009)*, 258–267.

Kersting, K.; Ahmadi, B.; and Natarajan, S. 2009. Counting Belief Propagation. In *Proc. of the 25th Conf. on Uncertainty in Artificial Intelligence (UAI–09)*.

Kersting, K. 2012. Lifted probabilistic inference. In *Proceedings of ECAI-2012*. IOS Press.

Niepert, M. 2012. Markov chains on orbits of permutation groups. In *Proc. of the 28th Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Ramana, M.; Scheinerman, E.; and Ullman, D. 1994. Fractional isomorphism of graphs. *Discrete Mathematics* 132:247–265.

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

Singla, P., and Domingos, P. 2008. Lifted First-Order Belief Propagation. In *Proc. of the 23rd AAAI Conf. on Artificial Intelligence (AAAI-08)*, 1094–1099.

Sontag, D., and Jaakkola, T. 2007. New outer bounds on the marginal polytope. In *Proc. of the 21st Annual Conference on Neural Information Processing Systems (NIPS-2007)*.

Sontag, D.; Globerson, A.; and Jaakkola, T. 2008. Clusters and coarse partitions in LP relaxations. In *Proc. of the 22nd Annual Conf. on Neural Information Processing Systems (NIPS)*, 1537–1544.

Sontag, D. 2010. *Approximate Inference in Graphical Models using LP Relaxations*. Ph.D. Dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.