Domain-Liftability of Relational Marginal Polytopes

Ondřej Kuželka
Czech Technical University in Prague, Czech Republic

Yuyi Wang
ETH Zurich, Switzerland

Abstract

We study computational aspects of *relational marginal polytopes* which are statistical relational learning counterparts of marginal polytopes, well-known from probabilistic graphical models. Here, given some first-order logic formula, we can define its relational marginal statistic to be the fraction of groundings that make this formula true in a given possible world. For a list of first-order logic formulas, the *relational marginal polytope* is the set of all points that correspond to expected values of the relational marginal statistics that are realizable. In this paper we study the following two problems: (i) Do domain-liftability results for the partition functions of Markov logic networks (MLNs) carry over to the problem of relational marginal polytope construction? (ii) Is the relational marginal polytope containment problem hard under some plausible complexity-theoretic assumptions? Our positive results have consequences for lifted weight learning of MLNs. In particular, we show that weight learning of MLNs is domain-liftable whenever the computation of the partition function of the respective MLNs is domain-liftable (this result has not been rigorously proven before).

1 Introduction

In this paper we study two problems about objects, called *relational marginal polytopes* [1], which arise naturally in the study of statistical relational learning models such as Markov logic networks [2]. A Markov logic network is given by a set of first-order logic formulas and their weights. Roughly speaking, the corresponding relational marginal polytope then represents the set

of all possible values of vectors of sufficient statistics that the Markov logic network with given first-order logic formulas may represent. Here, the sufficient statistics of the Markov logic network count the fraction of groundings of the first-order logic formulas defining it that are satisfied in a given possible world. For instance, for an MLN defined by formulas $\alpha_1 = sm(x)$ and $\alpha_2 = sm(x) \wedge friends(x, y) \wedge sm(y)$, the sufficient statistics would be the proportion of smokers in the population (the formula α_1) and the fraction of pairs of people who are friends and, at the same time, both smoke (the formula α_2).

The main advantages of specifying MLNs using sufficient statistics over specifying them using weights of formulas are (i) that it is clear what the parameters mean and (ii) that there are no issues with varying domain size caused by non-projectivity [3] of MLNs; we refer to [1, 4, 5] for further justification. An obvious disadvantage of specifying MLNs using sufficient statistics directly (instead of weights) is that they may be inconsistent, i.e. there may be no distribution satisfying them. Deciding if a given point lies inside the relational marginal polytope is equivalent to asking if there exists a distribution that has given expected values of the sufficient statistics.

Marginal polytopes are important objects which have been studied for the propositional case in the literature [6, 7, 8]. We believe it is important to generalize the most important results from non-relational PGM literature to the relational setting and we believe the study of *relational* marginal polytopes is one of them. It is well known that even just deciding whether a point is contained in a marginal polytope is NP-hard. A marginal polytope may have exponentially many vertices in the dimension of the modelled data. However, in the relational setting, in particular for Markov logic networks, the situation is more complex because there is no single notion of dimensionality of the data; instead there is the number of first-order logic formulas defining the Markov logic network and the size of the domain, which is the set of objects modelled by the Markov logic network (e.g. users of a social network). It turns out that the number of vertices of a *relational*

marginal polytope grows only polynomially with the size of the domain. It is quite frequent in applications of SRL that the number of first-order logic formulas is quite small and the domain is large, which is also one of the motivations for so-called *lifted inference* algorithms [9, 10, 11] which exploit symmetries stemming from the first-order nature of statistical relational models. It is therefore important to understand the properties of relational marginal polytopes, in particular the complexity of the relevant problems as a function of the domain size, which is what we study in this paper. In particular the main contributions of this paper are as follows: (i) We show that if a class of Markov logic networks admits a domain-lifted inference algorithm (in particular an algorithm for computing the partition function of Markov logic networks, with runtime polynomial in the size of the domain) then there is also a domain-lifted algorithm for constructing the respective relational marginal polytopes. This allows us to extend recent domain-liftability results for weight learning of Markov logic networks from [12]. (ii) We show that the problem of deciding whether a point is contained in a relational marginal polytope given by a set of first-order logic formulas and a domain of size n is, in general, not domain-liftable. This negative result for containment then also directly translates into a negative result for relational marginal polytope construction (because the number of vertices of the polytope is polynomial in the domain size).

The results in this paper are mostly meant as showing what is and what is not possible *theoretically*. Making the results *practical* is a different question.

2 Preliminaries

2.1 First Order Logic

We assume a function-free first-order language defined by a set of constants Δ, a set of variables $\mathcal{V}$ and for each $k \in \mathbb{N}$ a set $\mathcal{R}_k$ of k-ary predicates. Variables start with lowercase letters and constants start with uppercase letters. An atom is $r(a_1, ..., a_k)$ with $a_1, ..., a_k \in \Delta \cup \mathcal{V}$ and $r \in \mathcal{R}_k$. A literal is an atom or its negation. A clause is a universally quantified disjunction of a finite set of literals. A clause in which none of the literals contains any variables is called *ground*. The set of grounding substitutions of a clause α w.r.t. a set of constants Δ is the set $\Theta(\alpha, \Delta) = \{\vartheta_1, ..., \vartheta_m\}$ that contains substitutions to all variables occurring in α using constants from Δ. A possible world ω is represented as a set of ground atoms that are true in ω. The satisfaction relation $\models$ is defined in the usual way: $\omega \models \alpha$ means that the formula α is true in ω. When $\mathbf{x}$ is a list of first-order logic variables then $|\mathbf{x}|$ is used to denote the length of this list.

An important concept that we will need in this paper is the *spectrum of a first-order logic sentence*.

Definition 1 (Spectrum of sentence). *Let Φ be a first-order logic sentence. Its spectrum is the set of integers n for which Φ has a model on domain of size n.*

2.2 Complexity Classes ETIME and NETIME

In this section we briefly describe the necessary results from complexity theory that we will use for proving negative results in this paper.

Definition 2 (Complexity classes). *Given a function f, the class* **DTIME**$[f(n)]$ *(or* **NTIME**$[f(n)]$*) of decision problems are solvable by $f(n)$ time-bounded deterministic (or non-deterministic) Turing machines. The class* **ETIME** *is defined as* $\bigcup_{c>0}$ **DTIME**(2^{cn}) *and the class* **NETIME** *equals to* $\bigcup_{c>0}$ **NTIME**(2^{cn}).

The class **NETIME** is tightly related to spectra of first-order logic sentences via the following result of Jones and Selman.

Theorem 1 (Jones and Selman [13]). *A set $\mathcal{A} \subseteq \mathbb{N}$ is in* **NETIME** *if and only if $\mathcal{A}$ is the spectrum of a function-free and constant-free sentence Φ.*

2.3 Markov Logic Networks

A Markov logic network [2] (MLN) is a set of weighted first-order logic formulas (α, w), where $w \in \mathbb{R}$ and α is a function-free and quantifier-free first-order formula. The semantics are defined w.r.t. the groundings of the first-order formulas, relative to some finite set of constants Δ, called the domain. An MLN Φ induces the probability distribution over possible worlds $\omega \in \Omega$: $p_\Phi(\omega) = \frac{1}{Z} \exp\left(\sum_{(\alpha,w)\in\Phi} w \cdot N(\alpha, \omega)\right)$, where $N(\alpha, \omega)$ is the number of groundings of α satisfied in ω, and Z, called *partition function*, is a normalization constant to ensure that p_Φ is a probability distribution. It turns out to be more convenient to replace $N(\alpha, \omega)$ in the definition of MLNs by $Q(\alpha, \omega) = \frac{1}{|\Delta|^{|vars(\alpha)|}} \sum_{\vartheta \in \Theta(\alpha,\Delta)} \mathbb{1}(\omega \models \alpha\vartheta)$, where $\Theta(\alpha, \Delta)$ is the set of all grounding substitutions of α's variables using constants from Δ and $\mathbb{1}(\omega \models \alpha\vartheta)$ is the indicator function, which is equal to 1 when $\alpha\vartheta$ is true in the possible world ω. Thus, $Q(\alpha, \omega)$ is the fraction of the groundings of α satisfied in ω. Equivalently, we can say that $Q(\alpha, \omega)$ is the probability that $\omega \models \alpha\varphi$ if we pick φ from $\Theta(\alpha, \Delta)$ uniformly at random. With this notation we will write the probability of a possible world $\omega \in \Omega$ as: $p_\Phi(\omega) = \frac{1}{Z} \exp\left(\sum_{(\alpha,w)\in\Phi} w \cdot Q(\alpha, \omega)\right)$.

2.4 Relational Marginal Problems

An alternative way to view a Markov logic network $\Phi = \{(\alpha_1, w_1), \ldots, (\alpha_m, w_m)\}$ is to think of it as a maximum entropy distribution satisfying given marginal constraints $\mathbb{E}[Q(\alpha_i, .)] = \theta_i$. Assuming we have the expected values of the formula statistics $\mathbb{E}[Q(\alpha, .)]$, we can define the following maximum entropy problem [1].

Relational Marginal Problem (Formulation):

$$\min_{\{P_\omega \, : \, \omega \in \Omega\}} \sum_{\omega \in \Omega} P_\omega \log P_\omega \quad s.t. \quad (1)$$

$$\forall i = 1, \ldots, m : \sum_{\omega \in \Omega} P_\omega \cdot Q(\alpha_i, \omega) = \theta_i \quad (2)$$

$$\forall \omega \in \Omega : P_\omega \geq 0, \sum_{\omega \in \Omega} P_\omega = 1 \quad (3)$$

Here, the P_ω's are the problem's decision variables, each of which represents the probability of one possible world $\omega \in \Omega$. Line (1) is the maximum entropy criterion, which is shown here as the minimization of the negative entropy; Line (2) shows the constraints given by the statistics; and Line (3) provides the normalization constraints for the probability distribution.

Assuming there exists a feasible solution satisfying $\forall \omega : P_\omega > 0$ (we call such a solution *positive*), the optimal solution of the above maximum entropy problem is an MLN $P_\omega = \frac{1}{Z} \exp \left(\sum_{(\alpha_i, w_i) \in \Phi} w_i \cdot Q(\alpha_i, \omega) \right)$ where the parameters $\mathbf{w} = (w_1, \ldots, w_m)$ are obtained by maximizing the dual criterion, which also happens to be equivalent to the log-likelihood of the MLN w.r.t. a (possibly fictitious) training example $\widehat{\omega}$ that is over the same domain Δ and satisfies $Q(\alpha_i, \widehat{\omega}) = \theta_i$ for all the formula statistics (we refer to [1] for details).

2.5 Inference Using Weighted Model Counting

For exact inference in MLNs, one generally needs to be able to compute the partition function $Z = \sum_{\omega \in \Omega} e^{\sum_{\alpha_j} \lambda_j Q_\omega(\alpha_j)}$. Computation of the partition function Z can be converted to a *weighted first-order model counting problem* (WFOMC).

Definition 3 (WFOMC [11])**.** *Let $w(P)$ and $\overline{w}(P)$ be functions from predicates to real numbers (we call w and $\overline{w}$ weight functions) and let Γ be a first-order theory. Then $\mathbf{WFOMC}(\Phi, w, \overline{w}) = \sum_{\omega \in \Omega : \omega \models \Gamma} \prod_{a \in \mathcal{P}(\omega)} w(Pred(a)) \cdot \prod_{a \in \mathcal{N}(\omega)} \overline{w}(Pred(a))$, where $\mathcal{P}(\omega)$ and $\mathcal{N}(\omega)$ denote the positive literals that are true and false in ω, respectively, and $Pred(a)$ denotes the predicate of a (e.g. $Pred(friends(Alice, Bob)) = friends$).*

To compute the partition function Z using weighted model counting, we may proceed as in [11]. Let a Markov logic network Φ be given. Here, for simplicity of exposition, we will assume that the formulas in Φ do not contain constants (we refer to [11] for the general case). For every weighted formula $(\alpha_i, \lambda_i) \in \Phi$, where the free variables in α_i are exactly $x_1, \ldots, x_k$, we create a new formula $\forall x_1, \ldots, x_k : \xi_i(x_1, \ldots, x_k) \Leftrightarrow \alpha_i(x_1, \ldots, x_k)$ where ξ is a new fresh predicate; we denote the resulting set of formulas Γ. Then we set $w(\xi_i) = \exp \left(|\Delta|^{-|Vars(\alpha_i)|} \cdot \lambda_i \right)$ and $\overline{w}(\xi_i) = 1$ and for all other predicates we set both w and $\overline{w}$ equal to 1. It is easy to check that then $\mathbf{WFOMC}(\Gamma, w, \overline{w}) = Z$, which is what we needed to compute.

2.6 Liftability

Importantly, there are classes of first-order logic theories for which weighted model counting is polynomial-time. In particular, as shown in [14], when the theory consists only of first-order logic sentences, each of which contains at most two logic variables, the weighted model count can be computed in time polynomial in the number of elements in the domain Δ over which the set of possible worlds Ω is defined. It follows from the translation described in the previous section that this also means that computing the partition function of 2-variable MLNs can be done in time polynomial in the size of the domain and in the absolute value of the MLN's weights. This is not the case in general when the number of variables in the formulas is greater than two unless P = #P$_1$[1] [15]. Within statistical relational learning, the term used for problems that have such polynomial-time algorithms is *domain liftability*. Next we define what we mean by *domain liftability* [11].

Definition 4 (Domain liftability)**.** *An algorithm for computing the partition function Z of an MLN $\Phi = \{(\alpha_1, \lambda_1), \ldots, (\alpha_l, \lambda_l)\}$, where each λ_i is represented by two numbers[2] $a_i, b_i \in \mathbb{N}$ as $\lambda_i = \ln a_i - \ln b_i$, is said to be* domain-lifted *if it runs in time polynomial in the size of the domain Δ and in the number of bits needed to encode the numbers a_i and b_i. A class of MLNs is said to be* domain-liftable *if there is a domain lifted algorithm for computing the partition function Z for MLNs from this class.*

The definition that we use here differs slightly from the original definition by Van den Broeck [11] in that it also requires lifted algorithms to depend polynomially on

---

[1]#P$_1$ is the set of #P problems over a unary alphabet.

[2]The restriction on the representation of the weights ensures that the partition function will always be a rational number. Moreover, one can verify that the number of bits needed to represent the partition function will also be polynomial in the number of bits needed to represent the numbers a_i, b_i and in the domain size $|\Delta|$.

the size of the representation of the formulas' weights. A justification for this alternative definition follows from the work of Jaeger [16] (Section 4.2). In particular, all existing domain-lifted inference algorithms are also domain-lifted according to our definition. Another small technical difference is that we define domain-liftability directly in terms of complexity of computing the partition function Z.

3 Relational Marginal Polytopes

Now we can finally define relational marginal polytopes [1] which are the main focus of this paper. These represent the expected values for the vectors of statistics of the given formulas that are possible.

Definition 5 (Relational marginal polytope [1]). *Let Ω be the set of possible worlds on domain Δ and $\Phi = (\alpha_1, \ldots, \alpha_m)$ be a list of formulas. We define the relational marginal polytope $\mathbf{RMP}(\Phi, \Delta)$ w.r.t. Φ as $\mathbf{RMP}(\Phi, \Delta) = \{(x_1, \ldots, x_m) \in \mathbb{R}^m : \exists \text{ dist. on } \Omega \text{ s.t. } \mathbb{E}[Q(\alpha_1, \omega)] = x_1 \wedge \cdots \wedge \mathbb{E}[Q(\alpha_m, \omega)] = x_m\}$.*

The relational marginal polytope w.r.t. a given list of formulas $(\alpha_1, \ldots, \alpha_m)$ can be equivalently defined as the convex hull of the set $\{(Q(\alpha_1, \omega), \ldots, Q(\alpha_m, \omega)) : \omega \in \Omega\}$.

Next we illustrate the fact that relational marginal polytopes can be quite complex using the following example from [12]. In fact, we show later in this paper that, in general, the problem of deciding whether a point belongs to a relational marginal polytope is not a domain-liftable problem.

Example 2. *Consider the formulas $\alpha = friends(x_1, x_2)$ and $\beta = friends(x_1, x_2) \wedge friends(x_2, x_3) \wedge friends(x_3, x_1)$. Let $\Delta = \{C_1, \ldots, C_{100}\}$ be the set of domain elements and Ω be the respective set of possible worlds over the first-order language given by the predicate friends/2 and the constants from Δ. The possible worlds $\omega \in \Omega$ may be thought of as representing social networks. Then $Q(\alpha, \omega)$ corresponds to the "frequency" of friendships in the network and $Q(\beta, \omega)$ to the "frequency" of friendship-triangles. We can then see easily why there is, for instance, no distribution with $\mathbb{E}[Q(\alpha, \omega)] = 0$ and $\mathbb{E}[Q(\beta, \omega)] = 0.5$ (as graphs without edges cannot have a positive number of triangles). Hence, the point $(0, 1)$ will not be contained in the relational marginal polytope. In general relational marginal polytopes may be quite complicated objects as, for instance, in this case to construct the respective polytope we would also need to also find the "extremal" directed graphs with maximum number of triangles with a constrained number of edges etc.*

Next we define what it means for a point to be in the η-interior of a polytope.

Definition 6 (Interiority). *Let $\eta > 0$, $\mathbf{P}$ be a polytope and $A^= \mathbf{x} = \mathbf{c}$ be the maximal linearly independent system of linear equations that hold for the vertices of $\mathbf{P}$. A point θ is said to be in the η-interior of $\mathbf{P}$ if $\{\theta' | A^= \theta' = \mathbf{c}, \|\theta' - \theta\|_2 \leq \eta\} \subseteq \mathbf{P}$.*

We need to consider the system of linear equations $A^= \mathbf{x} = \mathbf{c}$ in the definition of interiority because the polytope may live in a lower dimensional subset of the given space. Our definition of interiority is also often called *relative interiority* in the literature.

4 A Lifted Reduction to Partition Functions of MLNs

In this section, we show that computing relational marginal polytopes can be reduced to computing the partition functions of some MLNs. Instead of $Q(\alpha, \omega)$, we deal with $N(\alpha, \omega)$ in this section, since there is a simple one-to-one mapping $Q(\alpha, \omega) = |\Delta|^{-|vars(\alpha)|} \cdot N(\alpha, \omega)$ and $N(\alpha, \omega)$ only uses integers. We call the convex hull of the set $\{(N(\alpha_1, \omega), \ldots, N(\alpha_m, \omega)) : \omega \in \Omega\}$ the *integer relational marginal polytope* $\mathbf{IRMP}(\Phi, \Delta)$. Let $P_i := \{N(\alpha_i, \omega) : \omega \in \Omega\}$. We can see that $\{(N(\alpha_1, \omega), \ldots, N(\alpha_m, \omega)) : \omega \in \Omega\}$ is a subset of $\bigtimes_{i=1}^{m} P_i$. The algorithm described in the following theorem is in polynomial time. Even though it would not be efficient in practice, our goal here is to show a reduction, instead of devise a fast algorithm.

Theorem 3. *Let $\Phi = (\alpha_1, \ldots, \alpha_m)$ be a list of first-order logic formulas. If the class of MLNs $\{(\alpha_i, \lambda_i)\}$ is domain-liftable then computing the facets of the relational marginal polytope $\mathbf{RMP}(\Phi, \Delta)$ can be done in time polynomial in $|\Delta|$.*

Proof. We first observe that, for any world ω, $N(\alpha_i, \omega)$ can only take value in $[r_i] = \{0, 1, \ldots, r_i\}$ where $r_i := \max P_i = O(|\Delta|^{|vars(\alpha_i)|})$. Hence, $N(\Phi, \omega)$ can only take value in $\bigtimes_{i=1}^{m}[r_i]$, which are integer points (vectors) in an m-dimensional space, whose i-th entries represent $N(\alpha_i, \omega)$.

If $\mathbf{IRMP}(\Phi, \Delta)$ is full-dimensional, each of its facets corresponds to a halfspace, which is of the form $\sum_i a_i x_i \leq b$, where x_i is the i-th coordinate, and can be written as $\vec{a} \cdot \vec{x} \leq b$. If $\mathbf{IRMP}(\Phi, \Delta)$ is not full-dimensional, each of its facets corresponds to an intersection of such half-spaces. To determine these halfspaces, we first list all possible normal vectors $\vec{a}$, and then find b for each normal vector.

Enumerate all possible $\vec{a}$'s Note that every vertex of the $\mathbf{IRMP}$ is an integer vector in $\bigtimes_{i=1}^{m}[r_i]$.

We list all possible linearly-independent m-tuples $(p^{(1)}, p^{(2)}, \ldots, p^{(m)})$ of $\bigtimes_{i=1}^{m}[r_i]$, where every $p^{(j)}$ is a point in $\bigtimes_{i=1}^{m}[r_i]$. For every tuple, we can efficiently compute the perpendicular vector of the hyperplane that passes every point in the tuple. For example, we subtract the coordinates of one of the points from all of the others and then compute their generalized cross product to obtain the perpendicular vector $\vec{v}$. To enumerate every possible $\vec{a}$, both $\vec{v}$ and $-\vec{v}$ need to be taken into account. In this way, every coordinate of every $\vec{a}$ is an integer (which is not necessary, but makes the next step easier). It is not difficult to see that the number of $\vec{a}$'s that we collect is polynomial in $|\Delta|$ and every entry a_i is polynomial in $|\Delta|$ as well.

Find b for each $\vec{a}$ For every vector $\vec{a}$, to determine the integer b for the halfspace $\sum_i a_i x_i \leq b$, we construct a Markov logic network $\{(\alpha_i, 2a_i \ln |\Omega|) : 1 \leq i \leq m\}$ and then compute the partition function of this MLN. If for every possible world ω, $\sum_i a_i N(\alpha_i, \omega) \leq b$, then the partition function is $\leq |\Omega|^{2b+1}$ since there are $|\Omega|$ possible worlds. If there exists a world ω such that $\sum_i a_i N(\alpha_i, \omega) > b$, then the partition function $\geq |\Omega|^{2b+2}$, since every a_i is an integer. Therefore, one can easily find the smallest integer b such that there does not exist any world satisfying $\sum_i a_i N(\alpha_i, \omega) > b$.

After the steps above, we obtain a set of linear inequality constraints (halfspaces). We can apply Lemma 2 (in the appendix) to get a minimal set of constraints that correspond to the facets of the **IRMP**, which can be transformed into the facets of **RMP** efficiently. $\square$

Here we would like to note that the techniques for lifted linear programming developed in the work [17, 18] are not directly relevant to our work. In our case, all the involved linear programs are already small (are polynomial in the size of the domain).

4.1 A Corollary for Lifted Weight Learning of Markov Logic Networks

Recently, it has been shown in [12] that weight learning of Markov logic networks based on maximization of log-likelihood is domain-liftable for the 2-variable fragment of Markov logic networks. Previously, it had been shown in [19] that computing the gradients of log-likelihood is domain-liftable for liftable fragments of Markov logic networks. What the work [12] added to this was to show that the whole weight learning problem is also domain-liftable, not just the procedure that computes the gradients. This is formally stated in the next theorem from [12].

Theorem 4 (Theorem 11 in [12])**.** *Let* $\Phi = \{\alpha_1, \ldots, \alpha_l\}$ *be a set of quantifier-free first-order logic formulas, each with at most 2 variables. Let Φ_0 be a*

set of universally quantified first-order logic sentences, each also with at most 2 variables. Let Ω_{Φ_0} be the set of models of Φ_0 over a given domain Δ. Let $\widehat{\omega} \in \Omega$ be a training example. Then there is an algorithm which finds weights of the MLN $\mathcal{M}$ given by formulas Φ such that the log-likelihood of $\mathcal{M}$ given the training example $\widehat{\omega}$ is within ε of the optimum. The algorithm runs in time polynomial in $|\Delta|$, $1/\varepsilon$ and $1/\eta$ where η is the interiority of the vector $Q_{\widehat{\omega}}(\Phi)$ in the relational marginal polytope $RMP(\Phi, \Omega_{\Phi_0})$.

The above theorem is restricted only to the 2-variable fragment of Markov logic networks but a bit larger tractable fragments were also studied in the literature, e.g. [20]. The only reason why the proof of Theorem 4 needs the assumption that the Markov logic network is from the 2-variable fragment is because the weight learning algorithm needs the respective relational marginal polytope and the work [12] only gives a domain-lifted algorithm for computing relational marginal polytopes for the 2-variable fragment. However, the result about domain-liftability that we presented in this section guarantees that construction of relational marginal polytopes is always domain-liftable for Markov logic networks with domain-liftable inference. Hence we get a strengthening of the results from [12].

Corollary 5. *The positive result stated in Theorem 4 holds for all Markov logic networks for which computing the partition function Z is domain-liftable, i.e. not just for the 2-variable fragment.*

5 Complexity of the Containment Test

In this section we study computational complexity of deciding if a point is contained in the relational marginal polytope given by a fixed list of first-order logic formulas and a domain Δ. In particular we are interested in the complexity w.r.t. the domain size as a parameter, i.e. in domain-liftability of the problem. We prove a negative result showing that, unless **ETIME = NETIME**, there is no algorithm for deciding if a point $\theta = (p_1/q_1, \ldots, p_k/q_k)$ is contained inside **RMP**(Φ, Δ) running in time polynomial in $|\Delta|$, $\sum_{i=1}^{k} p_i$ and $\sum_{i=1}^{k} q_i$.

The proof of the negative result borrows ideas from the work of Jaeger [16]. In particular, we will need the next two results.

Proposition 6 (Corollary 3.4 in [16])**.** *If* **ETIME $\neq$ NETIME** *then there exists a function-free and constant-free first-order sentence Φ such that $\{unary(n) | n \in spec(\Phi)\}$ is not recognized in deterministic polynomial time.*

Lemma 1 (Proposition 4.2 in [16])**.** *Let $\Phi(\mathbf{x})$ be a*

quantifier-free first-order logic formula (possibly containing constants and function symbols). Let $\mathcal{S}$ be the set of relation symbols and $\mathcal{S}^F$ the set of function symbols and constants contained in $\Phi(\mathbf{x})$. Let $\mathcal{S}^+$ be a set of new relation symbols that for every k-ary $f \in \mathcal{S}^F$ contains a $k+1$-ary R^f (constant symbols are treated as 0-ary function symbols). Let Func be a set of sentences such that for every $f \in \mathcal{S}^F$ it contains the following first-order logic sentences:

$$\forall \mathbf{x}, y, y' : R^f(\mathbf{x}, y) \wedge R^f(\mathbf{x}, y') \Rightarrow y = y' \qquad (4)$$

$$\forall \mathbf{x} \exists y : R^f(\mathbf{x}, y). \qquad (5)$$

Then there exists a formula $\Phi^+(\mathbf{x}, \mathbf{z})$ without quantifiers, function symbols and constants such that the following are equivalent for all n:

1. *there exists an $\omega \in \Omega_{\mathcal{S} \cup \mathcal{S}^F}$ on domain $\Delta = \{1, \ldots, n\}$ such that $\omega \models \forall \mathbf{x} : \Phi(\mathbf{x})$,*

2. *there exists an $\omega^+ \in \Omega_{\mathcal{S} \cup \mathcal{S}^+}$ on domain $\Delta = \{1, \ldots, n\}$ such that $\omega^+ \models Func \wedge \forall \mathbf{x}, \mathbf{z} : \Phi(\mathbf{x}, \mathbf{z})$.*

Next we state and prove our main negative result.

Theorem 7. *There exists a list of first-order logic formulas $\Gamma = (\alpha_1, \ldots, \alpha_k)$ such that the following holds. If **ETIME** $\neq$ **NETIME** then there is no algorithm for deciding if a point $\theta = (p_1/q_1, \ldots, p_k/q_k)$, where $p_i, q_i \in \mathbb{N}$ is contained in the relational marginal polytope **RMP**(Γ, Δ) that runs in time polynomial the size of the domain Δ and in $\sum_{i=1}^k p_i$ and $\sum_{i=1}^k q_i$. The hardness result holds even if the point θ is guaranteed not to be equal to a vertex of the polytope.*

Proof idea. We use the techniques which were devised by Jaeger to prove lower bounds on complexity of weighted model counting. Specifically, Proposition 6 tells us that there exists a function-free first-order logic sentence Φ^* whose spectrum is not recognizable in time polynomial in the size of the domain (assuming **ETIME** $\neq$ **NETIME**). Lemma 1 allows us to convert that sentence to a relatively manageable form. The conversion is done by first Skolemizing the sentence Φ^* (which introduces function symbols) and then using Lemma 1 to get rid of the function symbols. Although Lemma 1 introduces new existential quantifiers, these may then appear only in (5), which we will be able to deal with. Then what we need to show is that if we could decide membership of points in relational marginal polytopes efficiently then we could also efficiently decide satisfiability of Φ^* on domains of given size. This is not completely straightforward because, first, the formulas in Γ cannot have quantifiers and, second, the points should not correspond to statistics of individual possible worlds so that we could still guarantee that these points do not correspond to vertices

of the polytopes. Hence, we need to select the point θ carefully; we will place it quite (polynomially) close to a potential vertex of the polytope. Finally, we have to show that the point θ is indeed inside the polytope for domain of size n if and only if the sentence Φ^* has a model on a domain of size n.

Proof. Let Φ^* be a first-order logic sentence with non-polynomial-time spectrum whose existence follows from Proposition 6. Let $\forall \mathbf{x} : \Phi(\mathbf{x})$ be a Skolemization of Φ^* and let $\Phi^{Skol}(\mathbf{x}) = Func \wedge \forall \mathbf{x}, \mathbf{z} : \Phi(\mathbf{x}, \mathbf{z})$ be constructed from $\Phi(\mathbf{x})$ as in Lemma 1. Let us write

$$Func = (\forall \mathbf{x}, y, y' : \beta_1(\mathbf{x}, y, y')) \wedge (\forall \mathbf{x} \exists y : \gamma_1(\mathbf{x}, y)) \wedge \ldots$$
$$\wedge (\forall \mathbf{x}, y, y' : \beta_l(\mathbf{x}, y, y')) \wedge (\forall \mathbf{x} \exists y : \gamma_l(\mathbf{x}, y))$$

where $\beta_i(\mathbf{x}, y, y') = R^{f_i}(\mathbf{x}, y) \wedge R^{f_i}(\mathbf{x}, y') \Rightarrow y = y'$ and $\gamma_i(\mathbf{x}, y) = R^{f_i}(\mathbf{x}, y)$. Finally we define $\Phi' = A_{\Phi'}(\mathbf{x}, \mathbf{z}) \wedge \Phi(\mathbf{x}, \mathbf{z})$, and for every $1 \leq i \leq l$ we define

$$\beta'_i = A_{\beta_i}(\mathbf{x}, y, y') \wedge (R^{f_i}(\mathbf{x}, y) \wedge R^{f_i}(\mathbf{x}, y') \Rightarrow y = y'),$$
$$\gamma'_i = A_{\gamma_i}(\mathbf{x}, y) \wedge R^{f_i}(\mathbf{x}, y).$$

where $A_{\Phi'}$, A_{β_i}'s and A_{γ_i}'s are auxiliary relations which will be useful in turn. Then we construct the list of formulas $\Gamma = (\Phi', \beta'_1, \gamma'_1, \ldots, \beta'_l, \gamma'_l)$. Next all we need to do is to select a point θ such that the respective polytope-membership query would allow us to decide if Φ^* has a model on a domain of size n. We define $\Delta = \{1, 2, \ldots, n\}$, $\varepsilon = \frac{1}{l} \cdot \frac{1}{2l+2} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2}} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+1}+1}$, and $\theta = \left(1 - \varepsilon, 1 - \varepsilon, \frac{1-\varepsilon}{|\Delta|}, \ldots, 1 - \varepsilon, \frac{1-\varepsilon}{|\Delta|}\right)$. Here we note that $1/\varepsilon$ is polynomial in the size of the domain and so are the representations of Δ, ε and θ. It also follows from the way ε is selected that θ cannot be a vertex of the polytope.

Next we show that $\theta \in \mathbf{RMP}(\Gamma, \Delta)$ if and only if Φ^* has a model on a domain of size n.

($\Rightarrow$) We will show that if $\theta \in \mathbf{RMP}(\Gamma, \Delta)$ then $\forall \mathbf{x} : \Phi^{Skol}(\mathbf{x})$ has a model on a domain of size n. Since, by Lemma 1, $\forall \mathbf{x} : \Phi^{Skol}(\mathbf{x})$ has such a model if and only if Φ^* has a model on a domain of size n, it will also follow that if $\theta \in \mathbf{RMP}(\Gamma, \Delta)$ then Φ^* has a model on a domain of size n.

First, using Carathéodory's theorem, we know that θ can be written as a convex combination:

$$\theta = \sum_{j=1}^{2l+2} a_j \cdot Q(\Phi, \omega_j). \qquad (6)$$

Now we want to show that at least one of the possible worlds ω_i in (6), which we denote as $\widetilde{\omega}$, must satisfy the following three conditions:

1. $Q(\Phi', \widetilde{\omega}) = 1$,

2. $Q(\beta_i', \widetilde{\omega}) = 1$ for all $1 \le i \le l$,

3. $Q(\gamma_i', \widetilde{\omega}) = \frac{1}{|\Delta|}$ for all $1 \le i \le l$.

Let us denote $\theta' = (1, 1, 1/|\Delta|, 1, 1/|\Delta|, \ldots, 1, 1/|\Delta|)$. Then the above conditions can be also written as $Q(\Gamma, \widetilde{\omega}) = \theta'$. We start by showing that the first two conditions hold. Suppose, for contradiction, that one of these two conditions does not hold. Then $\|\theta' - \theta\|_\infty \ge \frac{1}{2l+2} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2}}$ which follows from the fact that $a_j \ge \frac{1}{2l+2}$ must hold for at least one of the a_j's in (6). However, we also have $\|\theta' - \theta\|_\infty \le \varepsilon$ and $\varepsilon < \frac{1}{2l+2} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2}}$, hence we have a contradiction. Therefore there must be at least one $\widetilde{\omega}$ that satisfies the first two conditions and there may, in fact, be multiple such worlds. Hence, $\widetilde{\omega}$ is a possible world on a domain of size n which is a model of the first-order logic sentences $\forall \mathbf{x}, \mathbf{z} : \Phi(\mathbf{x}, \mathbf{z})$, $\forall \mathbf{x}, y, y' : \beta_1(\mathbf{x}, y, y')$, $\ldots, \forall \mathbf{x}, y, y' : \beta_l(\mathbf{x}, y, y')$.

To show that at least one possible world from (6) must satisfy all three conditions, we will need to strengthen the above argument. We define $S_1 = \sum_{j \in I_1} a_j$ where I_1 is the set of indices of those possible worlds ω_j from (6) that satisfy the first two conditions, and $S_2 = 1 - S_1$. Using similar reasoning as above, we must have $S_2 \cdot \frac{1}{l+1} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2}} \le \varepsilon$ as otherwise we would have $\|\theta - \theta'\|_\infty > \varepsilon$, which can be seen as follows. If one of the sentences $\forall \mathbf{x}, \mathbf{z} : \Phi(\mathbf{x}, \mathbf{z})$, $\forall \mathbf{x}, y, y' : \beta_1(\mathbf{x}, y, y')$, $\ldots, \forall \mathbf{x}, y, y' : \beta_l(\mathbf{x}, y, y')$ is violated then $\|Q(\Gamma, \omega_j) - \theta'\|_\infty \ge 1/|\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2}$. We then obtain the inequality when we realize that each of the possible worlds ω_j, where $j \in I_2$, must violate at least one of the $l + 1$ formulas.

Hence it follows from the above that the next inequality must hold

$$S_2 \le (l + 1) \cdot |\Delta|^{|\mathbf{x}|+|\mathbf{z}|+2} \cdot \varepsilon. \tag{7}$$

Next we can notice that if $Q(\beta_i', \widetilde{\omega}) = 1$ then it must also hold $Q(\gamma_i', \widetilde{\omega}) \le \frac{1}{|\Delta|}$. Suppose, for contradiction, that none of the possible worlds $\widetilde{\omega} \in \{\omega_j | j \in I_1\}$ satisfies the third condition. Then it must be the case for each of these possible worlds that $Q(\gamma_i', \widetilde{\omega}) \le \frac{|\Delta|^{|\mathbf{x}|}-1}{|\Delta|^{|\mathbf{x}|+1}}$ for at least one γ_i'. We need the following to hold for every γ_i' from Γ: $\sum_{j \in I_2} a_i \cdot Q(\gamma_i', \omega_j) + \sum_{j \in I_1} a_j \cdot Q(\gamma_i', \omega_j) = \frac{1-\varepsilon}{|\Delta|}$ and we know that $\sum_{j \in I_2} a_i \cdot Q(\gamma_i', \omega_j) \le S_2$ and $\sum_{j \in I_1} a_i \cdot Q(\gamma_i', \omega_j) \le (1 - S_2) \cdot \frac{|\Delta|^{|\mathbf{x}|}-1/l}{|\Delta|^{\mathbf{x}+1}}$ for at least one γ_i'. It follows that we need (at the very least) the inequality $S_2 + (1 - S_2) \cdot \frac{|\Delta|^{|\mathbf{x}|}-1/l}{|\Delta|^{|\mathbf{x}|+1}} \ge \frac{1-\varepsilon}{|\Delta|}$ to be true. After simple algebraic manipulations this yields the inequality

$$S_2 \ge \frac{1/l - \varepsilon \cdot |\Delta|^{|\mathbf{x}|}}{|\Delta|^{|\mathbf{x}|+1} - |\Delta|^{|\mathbf{x}|} + 1/l} \ge \frac{1/l - \varepsilon \cdot |\Delta|^{|\mathbf{x}|}}{|\Delta|^{|\mathbf{x}|+1} + 1} \tag{8}$$

Plugging ε into (7) yields

$$S_2 \le \frac{1}{2} \cdot \frac{1}{l} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+1} + 1} \tag{9}$$

whereas plugging ε into (8) yields

$$S_2 \ge \frac{\frac{1}{l} - \frac{1}{2l+2} \cdot \frac{1}{l} \cdot \frac{1}{|\Delta|^{|\mathbf{z}|+2}} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+1}+1}}{|\Delta|^{|\mathbf{x}|+1} + 1}$$
$$\ge \frac{3}{4} \cdot \frac{1}{l} \cdot \frac{1}{|\Delta|^{|\mathbf{x}|+1} + 1}.$$

So we have arrived at a contradiction. It follows that at least one of the $\widetilde{\omega}$ must simultaneously satisfy conditions 1, 2 and 3. We have thus shown that if $\theta \in \mathbf{RMP}(\Gamma, \Delta)$ then Φ^* has a model on a domain of size n which is what we needed to show.

($\Leftarrow$) If Φ^* has a model on a domain of size n then so does $\forall \mathbf{x} : \Phi^{Skol}(\mathbf{x})$. Hence, let ω^* be such a model of $\forall \mathbf{x} : \Phi^{Skol}(\mathbf{x})$ on a domain of size n. Let ω^+ be a possible world on the set of relations of ω^* extended by the $A_{\Phi'}$ and A_{β_i} relations. Let ω^+ agree with ω^* on all the relations from ω^* and let ω^+ satisfy $\omega^+ \models (\forall \mathbf{x}, \mathbf{z} : A_{\Phi'}(\mathbf{x}, \mathbf{z}))$, $\omega^+ \models (\forall \mathbf{x}, y, y' : A_{\beta_i}(\mathbf{x}, y, y'))$ and $\omega^+ \models (\forall \mathbf{x}, y : A_{\gamma_i}(\mathbf{x}, y))$, for all $1 \le i \le l$. Then it is not difficult to check that $Q(\Gamma, \omega^+) = \left(1, 1, \frac{1}{|\Delta|}, 1, \frac{1}{|\Delta|}, \ldots, 1, \frac{1}{|\Delta|}\right)$. Next let ω^- be a possible world constructed from ω^* in almost the same way as ω^+ but this time satisfying $\omega^- \models (\forall \mathbf{x}, \mathbf{z} : \neg A_{\Phi'}(\mathbf{x}, \mathbf{z}))$, $\omega^- \models (\forall \mathbf{x}, y, y' : \neg A_{\beta_i}(\mathbf{x}, y, y'))$ and $\omega^- \models (\forall \mathbf{x}, y : \neg A_{\gamma_i}(\mathbf{x}, y))$, for all $1 \le i \le l$. Then we have $Q(\Gamma, \omega^-) = (0, 0, 0, 0, 0, \ldots, 0, 0)$. Since, by definition, $Q(\Gamma, \omega^+) \in \mathbf{RMP}(\Gamma, \Delta)$ and $Q(\Gamma, \omega^-) \in \mathbf{RMP}(\Gamma, \Delta)$ and since θ is a convex combination of $Q(\Gamma, \omega^+)$ and $Q(\Gamma, \omega^-)$, it follows that $\theta \in \mathbf{RMP}(\Gamma, \Delta)$. $\square$

6 Conclusion

In this paper we studied complexity of problems related to relational marginal polytopes. As our first main contribution we proved that domain-liftability of computing the partition function of a Markov logic networks carries over to the problem of constructing relational marginal polytopes, which allowed us to extend positive results on domain-liftability for weight learning of Markov logic networks from [12]. As our second main contribution we showed hardness of deciding whether a point is contained in a relational marginal polytope, assuming a widely believed complexity-theoretic conjecture.

In this paper we were interested only in answering the theoretical questions of domain-liftability: what is and is not possible. The next step is to design algorithms that will also be efficient in practice.

References

[1] Ondřej Kuželka, Yuyi Wang, Jesse Davis, and Steven Schockaert. Relational marginal problems: Theory and estimation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018.

[2] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.

[3] Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. *Annals of statistics*, 41(2):508, 2013.

[4] Manfred Jaeger and Oliver Schulte. Inference, learning, and population size: Projectivity for SRL models. *CoRR*, abs/1807.00564, 2018.

[5] Happy Mittal, Ayush Bhardwaj, Vibhav Gogate, and Parag Singla. Domain-size aware markov logic networks. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019*, pages 3216–3224, 2019.

[6] David Sontag and Tommi S Jaakkola. New outer bounds on the marginal polytope. In *Advances in Neural Information Processing Systems*, pages 1393–1400, 2008.

[7] Tim Roughgarden and Michael Kearns. Marginals-to-models reducibility. In *Advances in Neural Information Processing Systems*, pages 1043–1051, 2013.

[8] Guy Bresler, David Gamarnik, and Devavrat Shah. Hardness of parameter estimation in graphical models. In *Advances in Neural Information Processing Systems*, pages 1062–1070, 2014.

[9] Rodrigo De Salvo Braz, Eyal Amir, and Dan Roth. Lifted first-order probabilistic inference. In *Proceedings of the 19th international joint conference on Artificial intelligence*, pages 1319–1325. Citeseer, 2005.

[10] Vibhav Gogate and Pedro Domingos. Probabilistic theorem proving. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 256–265. AUAI Press, 2011.

[11] Guy Van den Broeck. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Advances in Neural Information Processing Systems*, pages 1386–1394, 2011.

[12] Ondřej Kuželka and Vyacheslav Kungurtsev. Lifted weight learning of markov logic networks revisited. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS-19)*, 2019.

[13] Neil D Jones and Alan L Selman. Turing machines and the spectra of first-order formulas. *The Journal of Symbolic Logic*, 39(1):139–150, 1974.

[14] Guy Van den Broeck, Wannes Meert, and Adnan Darwiche. Skolemization for weighted first-order model counting. In *Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2014.

[15] Paul Beame, Guy Van den Broeck, Eric Gribkoff, and Dan Suciu. Symmetric weighted first-order model counting. In *Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 313–328. ACM, 2015.

[16] Manfred Jaeger. Lower complexity bounds for lifted inference. *TPLP*, 15(2):246–263, 2015.

[17] Martin Mladenov, Babak Ahmadi, and Kristian Kersting. Lifted linear programming. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012*, pages 788–797, 2012.

[18] Kristian Kersting, Martin Mladenov, and Pavel Tokmakov. Relational linear programming. *Artif. Intell.*, 244:188–216, 2017.

[19] Jan Van Haaren, Guy Van den Broeck, Wannes Meert, and Jesse Davis. Lifted generative learning of markov logic networks. *Machine Learning*, 103(1):27–55, 2016.

[20] Seyed Mehran Kazemi, Angelika Kimmig, Guy Van den Broeck, and David Poole. New liftable classes for first-order probabilistic inference. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 3117–3125, 2016.