Quantified Markov Logic Networks*

Víctor Gutiérrez-Basulto
Cardiff University, UK
gutierrezbasultov@cardiff.ac.uk

Jean Christoph Jung
University of Bremen, Germany
KU Leuven, Belgium
jeanjung@uni-bremen.de

Ondřej Kuželka
KU Leuven, Belgium
ondrej.kuzelka@kuleuven.be

Abstract

Markov Logic Networks (MLNs) are well-suited for expressing statistics such as *"with high probability a smoker knows another smoker"* but not for expressing statements such as *"there is a smoker who knows most other smokers"*, which is necessary for modeling, e.g. influencers in social networks. To overcome this shortcoming, we study *quantified MLNs* which generalize MLNs by introducing *statistical universal quantifiers*, allowing to express also the latter type of statistics in a principled way. Our main technical contribution is to show that the standard reasoning tasks in quantified MLNs, maximum a posteriori and marginal inference, can be reduced to their respective MLN counterparts in polynomial time.

1 Introduction

Markov Logic Networks (MLNs) (Richardson and Domingos 2006) extend first-order logic (FOL) with means to capture uncertainty. This is intuitively achieved by softening the meaning of FOL formulas by associating a weight to them, such that the higher the weight, the higher the probability of the formula to be satisfied. Indeed, MLNs provide a compact representation of large Markov Networks with repeated substructures. The FOL component of MLNs makes them particularly suitable to represent background knowledge of a wide variety of application domains. As a consequence, MLNs have been successfully used to model knowledge in domains such as natural language (Riedel and McCallum 2011; Venugopal et al. 2014), computer vision (Tran and Davis 2008) and social network analysis (Chen et al. 2013; Farasat et al. 2015).

The kind of statistical regularities (that hold for a given problem) encoded by an MLN, directly depends on the type of quantifiers available in the language. Since MLNs are based on FOL, they come equipped with the standard FOL quantifiers $\exists$ and $\forall$. However, it has been observed that the modeling capabilities of these quantifiers might not be appropriate for certain application scenarios that require a kind of quantification describing, for instance *most*, *few*, or *at least* k thresholds, for more details see (Farnadi et al. 2017; Milch et al. 2008) and references therein, and Sec. 8 below.

In the present paper we investigate *Quantified Markov Logic Networks (QMNLs)*, the extension of classical MLNs with a *statistical quantifier* $\forall^*$. Indeed, MLNs lack means to describe certain types of statistics, e.g., the proportion of people, that are maximally connected to others. This type of modeling capabilities might be useful, for instance, in social network analysis to model influencers. As we shall see, with the use of the $\forall^*$ quantifier QMLNs are able to express this type of statistics.

Formally, this is done as follows: instead of weighted formulas $(\varphi(x_1,\dots,x_k),w)$ from MLNs, we use *weighted quantified sentences*

$$(Q_1x_1,\dots,Q_kx_k : \varphi(x_1,\dots,x_k),w),$$

where the Q_i can be arbitrary quantifiers $\exists,\forall,\forall^*$, and φ is a classical FOL formula, that is, the only quantifiers in φ are $\exists,\forall$. The semantics is given in terms of maximization ($\exists$), minimization ($\forall$), and expectation ($\forall^*$). While the semantics of the former ones is as expected, it is important to note that the semantics for the newly introduced statistical quantifier $\forall^*$ corresponds to uniform sampling of grounding substitutions of the respective variables. As a consequence, standard MLNs essentially correspond to the fragment of QMLNs where all Q_i are $\forall^*$ since in this case, the respective statistic represents the probability that the formula $\varphi(x_1,\dots,x_k)$ is true in a given possible world after grounding the variables $x_1,\dots,x_k$ using a randomly sampled substitution.

As a concrete example in QMLNs using the universal statistical quantifiers $\forall^*$, we can now measure the proportion of the population that are smokers and are known by one particular smoker, who knows most other smokers:

$$(\exists x\forall^* y\colon (\mathsf{smoker}(x)\wedge\mathsf{knows}(x,y)\wedge\mathsf{smoker}(y)),10).$$

Besides the mentioned practical reasons, the study of QMLNs also has a strong theoretical motivation. Let us recall the relation of MLNs with quantifier-free formulas and max entropy models constrained by statistics based on the random substitution semantics (Bacchus et al. 1992; Schulte et al. 2014). MLNs correspond to the solution of the maximum entropy relational marginal problem, where the modeled statistics are of the form $(\varphi(x_1,\dots,x_k),p)$ with $\varphi(x_1,\dots,x_k)$ a quantifier-free FOL formula and p is the probability that a random tuple $(a_1,\dots,a_k)$ satisfies $\varphi(a_1,\dots,a_k)$ in the model that we learn from (Kuželka et

*Authors appear in strict alphabetical order.

al. 2018). Given that QMLNs have higher expressive power, it is interesting to investigate whether a similar correspondence also exists for QMLNs – of course given more expressive statistics. It is easy to see that previously used techniques lift to restricted QMLNs with quantifier prefix $\forall^*$. Here we show that the correspondence also holds for arbitrary QMLNs and the respective statistics.

Objective and Contributions The main objective of this paper is to introduce QMLNs as an extension of MLNs with means to express more complex statistics, and to develop technical foundations for these QMLNs. Our main technical contributions are (i) the establishment of basic properties of QMNLs, analogous to those existing for standard MLNs; (ii) a generalization of the random substitution semantics to QMLNs and (iii) a polynomial time translation from QMLNs to MLNs, yielding a polytime reduction of the maximum a posteriori and marginal inference in QMLNs to their respective variant problems in standard MLNs. (iv) Furthermore, we pinpoint certain implications of extending MLNs to QMLNs in the context of *symmetric weighted first-order model counting (WFOMC)*.

Outline of the Paper After providing some preliminaries in Section 2, we introduce the syntax and semantics of quantified Markov logic networks in Section 3. Then, in Section 4 basic results on the treatment of negative weights and weights tending to infinity are provided. We also present results on the duality of relational marginal problems and QMLNs. Sections 5 to 6 contain our translations from QMLNs to MLNs, establishing the polytime reductions described above. Section 7 discusses the relation between QMLNs restricted to two variables and WFOMC. Section 8 presents related work and Section 9 conclusions and future work.

2 Background and Notation

We next provide some basics on First-Order Logic, Markov Logic Networks and Relational Marginal Problems.

2.1 First-Order Logic (FOL)

We give a short review of the function-free fragment of first-order logic (FOL), considered in this paper. Let $\mathbf{C} = \{a, b, \ldots\}$ be a finite set of *constants* and $\mathbf{V} = \{x, y, \ldots\}$ an infinite set of *variables*. A *term* t is an element in $\mathbf{C} \cup \mathbf{V}$. An *atom* is an expression of the form $R(t_1, \ldots, t_n)$, where R is a *predicate name* with *arity* n and terms t_i. As usual, a FOL-*formula* φ is constructed from atoms using *logical connectives* $\neg, \wedge, \vee, \Rightarrow$ and *quantifiers* $\exists$ *and* $\forall$. We assume the reader is familiar with the standard notions of quantified and free variables, and sentence. Given a formula φ, a variable x and a constant a, we use $\varphi[x/a]$ to denote the result of substituting in φ every occurrence of x with a. Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{a} = (a_1, \ldots, a_n)$ be tuples of variables and constants, respectively, we write $\varphi[\mathbf{x}/\mathbf{a}]$ to denote the application of $\varphi[x_i/a_i]$ for all $1 \leq i \leq n$. The *grounding of a formula* $\varphi(\mathbf{x})$ *over a domain* Δ, denoted with $\mathsf{gr}(\varphi, \Delta)$, is the set of all possible sentences obtained from φ by substituting all its free variables $\mathbf{x} = (x_1, \ldots, x_k)$ with any possible combination of constants from Δ, that is,

$$\mathsf{gr}(\varphi, \Delta) = \{\varphi[\mathbf{x}/\mathbf{a}] \mid \mathbf{a} \in \Delta^k\}.$$

A *vocabulary* σ is a finite set of predicate names such that each predicate name $R \in \sigma$ is associated with an *arity* $\mathsf{ar}(R)$. Given a vocabulary σ and a domain Δ, a *σ-structure over* Δ is any set ω consisting only of facts of the form $R(a_1, \ldots, a_n)$ such that $R \in \sigma$, $\mathsf{ar}(R) = n$, and $a_1, \ldots, a_n \in \Delta$. We denote with $\Omega(\sigma, \Delta)$ the set of all σ-structures over Δ and refer to the members of $\Omega(\sigma, \Delta)$ with *possible worlds*. Throughout the paper we often do not make the vocabulary explicit; it is then assumed to be just the set of predicate names used in the formulas.

The semantics of FOL is defined as usual in the context of Markov Logic Networks. More precisely, we use only finite domains Δ and assume that always $\mathbf{C} \subseteq \Delta$. Note that we refrain from using a many-sorted domain for the sake of simplicity; all our results lift to that case. Formally, we write $\omega \models \varphi$ when a sentence φ is satisfied in a structure ω. Given a set Φ of sentences, we write $\omega \models \Phi$ if $\omega \models \varphi$ for all $\varphi \in \Phi$.

2.2 Markov Logic Networks (MLNs)

A *Markov Logic Network (MLN)* is a finite set of weighted formulas (φ, w), where $w \in \mathbb{R} \cup \{+\infty\}$ is a weight and φ is a FOL-formula, possibly with free variables. If $w = +\infty$ then (φ, w) is called a *hard constraint*, otherwise a *soft constraint*. As an example, the soft constraint

$$(\mathsf{smoker}(x) \wedge \mathsf{friends}(x, y) \Rightarrow \mathsf{smoker}(y), 10)$$

intuitively asserts that having friends who are smokers makes one more likely to be a smoker as well, given the weight 10 is positive.

The semantics of MLNs is defined following (Van den Broeck, Meert, and Darwiche 2014).[1] A given MLN $\Phi = \{(\varphi_1, w_1), (\varphi_2, w_2), \ldots, (\varphi_k, w_k)\}$ and a domain Δ describe a probability distribution $p_{\Phi,\Delta}$ over $\Omega(\sigma, \Delta)$. To define the distribution, let $\Phi_S \subseteq \Phi$ denote the set of soft constraints in Φ and Φ_H denote the set of FOL sentences φ' obtained from the the hard constraints $(\varphi, +\infty) \in \Phi$ by adding a prefix of universal quantifiers for all free variables in φ. For instance, if $\Phi = \{(\mathsf{smoker}(x), +\infty)\}$ then $\Phi_H = \{\forall x : \mathsf{smoker}(x)\}$. Now, the distribution is defined by taking, for every $\omega \in \Omega(\sigma, \Delta)$,

$$p_{\Phi,\Delta}(\omega) = \begin{cases} \frac{1}{Z} \exp\left(\sum_{(\varphi,w)\in\Phi_S} w \cdot N(\varphi, \omega)\right) & \omega \models \Phi_H \\ 0 & \text{otherw.} \end{cases}$$

where $N(\varphi, \omega)$ is the number of sentences $\varphi' \in \mathsf{gr}(\varphi, \Delta)$ such that $\omega \models \varphi'$, and Z is a normalization constant.

Reasoning Problems We study the following problems.

Maximum a posteriori (MAP):

- **Input:** an MLN Φ and a domain Δ

[1] This slightly differs from the semantics introduced in (Richardson and Domingos 2006), we discuss it in the related work section.

- **Problem:** determine the world $\omega \in \Omega(\sigma, \Delta)$ maximizing $p_{\Phi,\Delta}(\omega)$.

Marginal inference (MARG):

- **Input:** an MLN Φ, a domain Δ and a FOL sentence φ
- **Problem:** compute the probability $Pr_{\Phi,\Delta}(\varphi)$ of φ, i.e.

$$Pr_{\Phi,\Delta}(\varphi) = P_{\omega \sim p_{\Phi,\Delta}(\omega)}[\omega \models \varphi],$$

where a subscript $d \sim D$ refers to sampling d according to a distribution D.

2.3 Relational Marginal Problems

MLNs containing only *quantifier-free* FOL formulas can be seen as solutions to a maximum entropy problem (Kuželka et al. 2018) constrained by *statistics* which are based on the *random-substitution semantics* (Bacchus et al. 1992; Schulte et al. 2014). Such statistics are defined as follows. For a possible world $\omega \in \Omega(\sigma, \Delta)$ and a quantifier-free FOL formula $\varphi(\mathbf{x})$ with k free variables $\mathbf{x}$, the *statistic* $Q_\omega(\varphi(\mathbf{x}))$ *of* $\varphi(\mathbf{x})$ is defined as

$$Q_\omega(\varphi(\mathbf{x})) = \mathbb{E}_{\mathbf{a} \sim \mathrm{Unif}(\Delta^k)} \left[\mathbb{1} \left(\omega \models \varphi[\mathbf{x}/\mathbf{a}] \right) \right]$$

where $\mathrm{Unif}(\Delta^k)$ denotes the uniform distribution over elements of the set Δ^k.[2]

Intuitively, the statistics $Q_\omega(\varphi(\mathbf{x}))$ measures how likely it is that the formula $\varphi(\mathbf{x})$ is satisfied in ω when a random substitution of $\mathbf{x}$ by domain elements is picked. The statistics $Q_\omega(\varphi(\mathbf{x}))$ can then be straightforwardly extended to statistics of probability distributions $Q[\varphi(\mathbf{x})]$. For a distribution $p(\omega)$ over $\Omega(\sigma, \Delta)$, $Q[\varphi(\mathbf{x})]$ is defined as

$$Q[\varphi(\mathbf{x})] = \mathbb{E}_\omega \left[Q_\omega(\varphi(\mathbf{x})) \right] = \sum_{\omega \in \Omega(\sigma,\Delta)} p(\omega) \cdot Q_\omega(\varphi(\mathbf{x})).$$

Now, the *maximum entropy relational marginal problem* is defined as follows:

- **Input**: set of statistics $\{(\varphi_1(\mathbf{x}_1), q_1), \ldots, (\varphi_n(\mathbf{x}_n), q_n)\}$, domain Δ
- **Problem**: find a distribution $p^*(\omega)$ over $\Omega(\sigma, \Delta)$ which has maximum entropy and satisfies the constraints

$$Q[\varphi_1(\mathbf{x_1})] = q_1, \ldots, Q[\varphi_n(\mathbf{x}_n)] = q_n.$$

To motivate the problem, we note that it has been shown in (Kuželka et al. 2018) that

(i) the solution of the relational marginal problem is a Markov logic network of the form $\{(\varphi_1(\mathbf{x}_1), w_1), \ldots, (\varphi_n(\mathbf{x}_n), w_n)\}$ where the weights w_i are obtained from the dual problem of the maximum entropy relational marginal problem, and

(ii) if the input probabilities $q_1, \ldots, q_n$ are estimated from data with domain Δ' with $|\Delta'| = |\Delta|$, then the result of the maximum entropy relational marginal problem coincides with maximum likelihood estimation, the most common approach to weight learning in MLNs (Richardson and Domingos 2006).

[2]Here, we depart slightly from (Kuželka et al. 2018) in that we do not require the random substitutions to be injective. However, the duality of MLNs and relational marginal problems holds as well in the case of non-injective substitutions.

In a sense, Point (i) can be viewed as the relational generalization of what has been done in the propositional setting (Singh and Vishnoi 2014). Let us finally remark that Point (ii) does not hold when $|\Delta| \neq |\Delta'|$ which follows from the results by Shalizi and Rinaldo (2013). Hence, the relational marginal view is more general from the statistical point of view; we refer to (Kuželka et al. 2018) for details.

3 Quantified Markov Logic Networks

We introduce the notion of *Quantified Markov Logic Networks (QMLNs)*, a generalization of standard MLNs capable of expressing expectations using "statistical" quantifiers. In QMLNs, the main ingredients of MLNs – weighted formulas (φ, w) with w a weight and φ a FOL formula – are replaced with weighted quantified sentences (α, w).

Definition 1 (Quantified Sentence)**.** *A* quantified sentence *is a formula α with*

$$\alpha = Q_1 x_1 \ldots Q_n x_n \colon \varphi(x_1, \ldots, x_n),$$

where each Q_i is a quantifier from $\{\forall, \forall^*, \exists\}$ and $\varphi(x_1, \ldots, x_n)$ is a classical first-order formula with free variables precisely $x_1, \ldots, x_n$.

Note that every FOL sentence is also a quantified sentence, but conversely a quantified sentence using the quantifier $\forall^*$ is not a FOL sentence.

Definition 2 (QMLNs)**.** *A* Quantified Markov Logic Network (QMLN) *is a finite set Φ of pairs (α, w) such that α is a quantified sentence and $w \in \mathbb{R} \cup \{+\infty, -\infty\}$.*

Before we can give the semantics of QMLNs, we give the semantics for quantified sentences. Intuitively, given a quantified sentence α and a possible world ω, we measure the extent to which α is satisfied in ω.

Definition 3 (Sentence Statistics)**.** *Let $\omega \in \Omega(\sigma, \Delta)$ be a possible world and α be a quantified sentence. Then the α-statistic of ω, denoted $Q_\omega(\alpha)$, is defined as follows:*

- *if α is an FOL sentence, then*

$$Q_\omega(\alpha) = \mathbb{1} \left(\omega \models \alpha \right), \tag{1}$$

- *if $\alpha = \forall x \colon \alpha'(x)$ is not an FOL sentence, then*

$$Q_\omega(\alpha) = \min_{a \in \Delta} Q_\omega(\alpha'[x/a]), \tag{2}$$

- *if $\alpha = \exists x \colon \alpha'(x)$ is not an FOL sentence, then*

$$Q_\omega(\alpha) = \max_{a \in \Delta} Q_\omega(\alpha'[x/a]), \tag{3}$$

- *if $\alpha = \forall^* x \colon \alpha'(x)$ is not an FOL sentence, then*

$$Q_\omega(\alpha) = \frac{1}{|\Delta|} \sum_{a \in \Delta} Q_\omega(\alpha'[x/a]). \tag{4}$$

Note that the case (1) in the above definition is only applied when α is a classical FOL formula, that is, when it does not contain any $\forall^*$ quantifiers and, as such, it serves as a base case of the recursive definition. Moreover, the case of the quantifier $\forall^*$ in (4) above can be alternatively expressed as

$$Q_\omega(\alpha) = \mathbb{E}_{a \sim \mathrm{Unif}(\Delta)} \left[Q_\omega(\alpha'[x/a]) \right] \tag{5}$$

where the expectation is w.r.t. a uniform distribution of a over Δ. From this we see that the definition of statistics $Q_\omega(\alpha)$ given by Definition 3 generalizes that of statistics based on random substitution semantics, cf. Section 2.3.

Remark 1. *We can easily check the following property of sentence statistics. Let α be a sentence and $\omega \in \Omega(\sigma, \Delta)$ be a possible world. If α' is obtained from α by replacing every quantifier $\forall^*$ by its classical counterpart $\forall$ then*

$$Q_\omega(\alpha) = 1 \text{ iff } \omega \models \alpha'.$$

As a result of this, we will sometimes abuse notation and write $\omega \models \alpha$ when $Q_\omega(\alpha) = 1$ even if α is not an FOL sentence.

Example 1. In classical first-order logic, the sentence

$$\exists x \forall y \colon \mathsf{knows}(x, y)$$

asserts that there is someone who knows everyone else (e.g. in a social network). If we replace $\forall y$ by $\forall^* y$, we get a quantified sentence

$$\alpha = \exists x \forall^* y \colon \mathsf{knows}(x, y) \tag{6}$$

which relaxes the hard constraint. Indeed, its associated statistic $Q_\omega(\exists x \forall^* y \colon \mathsf{knows}(x, y))$ measures the maximal proportion of people known by a single domain element. In graph-theoretical terms, this corresponds to the maximum out-degree of domain elements. Note that we could not directly express the same statistics in normal MLNs since, using normal MLNs, we could only express statistics corresponding to the sentence $\forall^* x \exists y : \mathsf{knows}(x, y)$, which intuitively measures the proportion of people who know at least one person. As we show later in the paper, it is possible to express MLNs with constraints encoding the same statistics but in order to do that we will have to enlarge the vocabulary σ, introducing additional predicates. △

We now have almost all the ingredients to define the semantics of QMLNs. What remains is to extend the definition of Φ_H, i.e. the hard constraints. Given a QMLN $\Phi = \{(\alpha_1, w_1), (\alpha_2, w_2), \ldots, (\alpha_k, w_k)\}$, we define Φ_H for QMLNs as follows. First, we define $\Phi_H^{+\infty}$ to be the set of FOL sentences obtained from weighted sentences $(\alpha, +\infty) \in \Phi$ by replacing all $\forall^*$ quantifiers by the classical $\forall$ quantifiers. Second, we define $\Phi_H^{-\infty}$ to be the set of FOL sentences α' that are obtained from weighted sentences $(\alpha, -\infty) \in \Phi$ where $\alpha = Q_1 x_1, \ldots, Q_n x_n : \varphi(x_1, \ldots, x_n)$, as $\alpha' = \widetilde{Q}_1 x_1, \ldots, \widetilde{Q}_n x_n : \neg\varphi(x_1, \ldots, x_n)$ where $\widetilde{\forall^*} = \forall$, $\widetilde{\forall} = \exists$ and $\widetilde{\exists} = \forall$. Finally, we define $\Phi_H = \Phi_H^{-\infty} \cup \Phi_H^{+\infty}$.

Next we define the semantics of QMLNs.

Definition 4 (Semantics of QMLNs)**.** *Given a QMLN $\Phi = \{(\alpha_1, w_1), (\alpha_2, w_2), \ldots, (\alpha_k, w_k)\}$ and a domain Δ, the probability of a possible world $\omega \in \Omega(\sigma, \Delta)$ is defined as:*

$$p_{\Phi,\Delta}(\omega) = \begin{cases} \frac{1}{Z} \exp\left(\sum_{(\alpha,w)\in\Phi_S} w \cdot Q_\omega(\alpha)\right) & \omega \models \Phi_H \\ 0 & \text{otherwise} \end{cases}$$

where $Q_\omega(\alpha)$ is the α-statistic of ω, and Z is a normalization constant.

We illustrate the semantics by continuing Example 1.

Example 1 (continued). Consider again the quantified sentence α from Equation (6) in Example 1. If you include the weighted quantified sentence (α, w) for some $w > 0$ in a QMLN, worlds in which there is an individual who knows most of the people get a higher probability than worlds for which this is not the case. As a result, worlds that have an "influencer" (and are thus closer to a social network) are considered more likely. △

Definition 5 (Marginal query problem)**.** *Let α be a sentence and $p_{\Phi,\Delta}(\omega)$ be the probability distribution over $\Omega(\sigma, \Delta)$ induced by the QMLN Φ and domain Δ. The* marginal query problem *is to compute the marginal probability defined as:*

$$Q_{\Phi,\Delta}[\alpha] = \mathbb{E}_\omega\left[Q_\omega(\alpha)\right] = \sum_{\omega\in\Omega(\sigma,\Delta)} p_{\Phi,\Delta}(\omega) \cdot Q_\omega(\alpha).$$

Remark 2. *If α is a sentence that does not contain $\forall^*$ quantifiers and $p_{\Phi,\Delta}(\omega)$ is the induced distribution over $\Omega(\sigma, \Delta)$ then $Q_{\Phi,\Delta}[\alpha] = P_{\omega\sim p_{\Phi,\Delta}(\omega)}[\omega \models \alpha]$, that is, it coincides with the marginal probability of α. Thus, Definition 5 generalizes the classical definition of marginal inference.*

4 Initial Observations

The goal of this section is to make some initial observations about QMLNs, which will be either of independent interest or exploited in some later proof.

First, it is easy to see that QMLNs generalize MLNs in the sense that we can view a weighted formula (φ, w) as a quantified sentence with implicit $\forall^*$-quantification over all free variables of φ. More formally, we have:

Proposition 1. *Let Φ_0 be an MLN and obtain a QMLN Φ from Φ_0 by replacing every weighted formula $(\varphi(x_1, \ldots, x_n), w) \in \Phi_0$ with the weighted quantified sentence $(\forall^* x_1 \ldots \forall^* x_n \colon \varphi(x_1, \ldots, x_n), w \cdot |\Delta|^n)$. Then, for every Δ and $\omega \in \Omega(\sigma, \Delta)$, we have $p_{\Phi_0,\Delta}(\omega) = p_{\Phi,\Delta}(\omega)$.*

4.1 Negation in QMLNs

It is well-known that in classical MLNs it is without loss of generality to assume positive weights. We show an analogous property of QMLNs.

Definition 6 (Negation)**.** *We define the negation $\mathrm{neg}(\alpha)$ of quantified sentences $\alpha = Q_1 x_1 \ldots Q_n x_n \colon \varphi(x_1, \ldots, x_n)$ by taking*

$$\mathrm{neg}(\alpha) = \overline{Q}_1 x_1 \ldots \overline{Q}_n x_n \colon \neg\varphi(x_1, \ldots, x_n),$$

where $\overline{\exists}$ is $\forall$, $\overline{\forall}$ is $\exists$, and $\overline{\forall^*}$ is $\forall^*$.

It is easy to check that $\mathrm{neg}(\mathrm{neg}(\alpha)) = \alpha$. Next we illustrate the way negation works in our setting on a concrete example.

Example 2. Let us see what happens if we take the sentence $\exists x \forall^* y \colon \mathsf{knows}(x, y)$ from Example 1 and negate it. Using Definition 6, we obtain

$$\mathrm{neg}\left(\exists x \forall^* y \colon \mathsf{knows}(x, y)\right) = \forall x \forall^* y \colon \neg\mathsf{knows}(x, y).$$

For the statistic $Q_\omega\left(\forall x \forall^* y\colon \neg \mathsf{knows}(x,y)\right)$ we have

$$\begin{aligned}
&Q_\omega\left(\forall x \forall^* y\colon \neg \mathsf{knows}(x,y)\right)\\
&\quad= \min_{t\in\Delta} \frac{1}{|\Delta|} \sum_{u\in\Delta} \mathbb{1}\left(\omega \models \neg \mathsf{knows}(t,u)\right)\\
&\quad= \min_{t\in\Delta} \frac{1}{|\Delta|} \sum_{u\in\Delta} \left(1 - \mathbb{1}\left(\omega \models \mathsf{knows}(t,u)\right)\right)\\
&\quad= 1 - \max_{t\in\Delta} \frac{1}{|\Delta|} \sum_{u\in\Delta} \mathbb{1}\left(\omega \models \mathsf{knows}(t,u)\right)\\
&\quad= 1 - Q_\omega(\exists x \forall^* y\colon \mathsf{knows}(x,y)).
\end{aligned}$$

△

In Example 2, the statistic of a negation of a sentence α turns out to be equal to one minus the statistic of that sentence, which is intuitively desirable. By repeatedly applying the shown argument, one can show that this holds in general:

Proposition 2. *For any sentence α and any possible world $\omega \in \Omega(\sigma,\Delta)$ the following holds:*

$$Q_\omega(\mathrm{neg}\left(\alpha\right)) = 1 - Q_\omega(\alpha).$$

Next we show that the same distribution represented by a QMLN Φ can be represented by another QMLN in which we replace some of the sentences by their negations while also inverting the signs of their respective weights. To show that we will need the next lemma.

Lemma 3. *Let Δ be a finite domain and Φ a QMLN. Let α, β be two sentences and $\Phi_\alpha = \{(\alpha, w)\} \cup \Phi$, $\Phi_\beta = \{(\beta, w)\} \cup \Phi$ and $\Phi'_\beta = \{(\beta, -w)\} \cup \Phi$ where w is finite. Then*

1. *if $Q_\omega(\alpha) = Q_\omega(\beta) + C_\Delta$ for all $\omega \in \Omega(\sigma,\Delta)$ for some constant C_Δ, then $p_{\Phi_\alpha,\Delta}(\omega) = p_{\Phi_\beta,\Delta}(\omega)$.*
2. *if $Q_\omega(\alpha) = -Q_\omega(\beta) + C_\Delta$ for all $\omega \in \Omega(\sigma,\Delta)$ for some constant C_Δ, then $p_{\Phi_\alpha,\Delta}(\omega) = p_{\Phi'_\beta,\Delta}(\omega)$.*

Proof. For the first case, we have

$$\begin{aligned}
\frac{p_{\Phi_\alpha,\Delta}(\omega)}{p_{\Phi_\beta,\Delta}(\omega)} &= \frac{Z_\beta}{Z_\alpha} \cdot \frac{\exp\left(\sum_{(\alpha,w)\in\Phi_\alpha} w \cdot Q_\omega(\alpha_i)\right)}{\exp\left(\sum_{(\alpha',w')\in\Phi_\beta} w' \cdot Q_\omega(\alpha')\right)}\\
&= \frac{Z_\beta}{Z_\alpha} \exp\left(w \cdot C_\Delta\right)\\
&= \frac{\sum_{\omega'\in\Omega(\sigma,\Delta)} \exp\left(\sum_{(\alpha',w')\in\Phi_\beta} w' \cdot Q_{\omega'}(\alpha')\right)}{\sum_{\omega'\in\Omega(\sigma,\Delta)} \exp\left(\sum_{(\alpha,w)\in\Phi_\alpha} w \cdot Q_{\omega'}(\alpha)\right)} \cdot e^{w\cdot C_\Delta}\\
&= \frac{1}{e^{w\cdot C_\Delta}} \cdot e^{w\cdot C_\Delta} = 1
\end{aligned}$$

The reasoning for the second case, $Q_\omega(\alpha) = -Q_\omega(\beta) + C_\Delta$, is completely analogical. □

Proposition 4. *Let $\Phi = \{(\alpha_1, w_1), \ldots, (\alpha_k, w_k)\}$ and $\Phi' = \{(\mathrm{neg}\left(\alpha_1\right), -w_1), \ldots, (\alpha_k, w_k)\}$. Then, for every domain Δ, and every $\omega \in \Omega(\sigma,\Delta)$, we have:*

$$p_{\Phi,\Delta}(\omega) = p_{\Phi',\Delta}(\omega).$$

Proof. The proof follows straightforwardly from Lemma 3 above and Proposition 2 for finite weights and from the definition of semantics of QMLNs for infinite weights. □

It follows from Proposition 2 and Proposition 4 that we can focus on QMLNs that have only positive weights. It also follows that it makes no sense to have a sentence and its negation in the set of sentences defining an QMLN.

4.2 Limit $w \to \infty$

In the seminal paper on Markov logic networks (Richardson and Domingos 2006) it was shown that if the weights of formulas of an MLN tend to infinity at the same pace, in the limit the MLN will define a uniform distribution over models of the classical first-order logic theory consisting of the MLN's rules. More precisely, let us denote with $\widehat{\Phi}$ the first-order logic sentence obtained from a given $\Phi = \{(\varphi_1, w), \ldots, (\varphi_n, w)\}$ by taking the conjunction of all formulas of the shape $\forall x_1, \ldots, x_k\colon \varphi_i(x_1, \ldots, x_k)$ where $x_1, \ldots, x_k$ are precisely the free variables in φ_i. Then the possible worlds that have non-zero probability for $w \to \infty$ are precisely the models of $\widehat{\Phi}$. The next proposition generalizes this by establishing that an analogical property also holds for QMLNs.

Proposition 5. *Let Δ be a finite domain and $\Phi(w) = \{(\alpha_1, w), \ldots, (\alpha_n, w)\}$ be a QMLN where every weight is w, and let $\widehat{\Phi}$ denote the FOL sentence obtained from $\alpha_1 \wedge \ldots \wedge \alpha_n$ by replacing every occurrence of $\forall^*$ by $\forall$. If $\widehat{\Phi}$ has a model in $\Omega(\sigma,\Delta)$, then, for every $\omega \in \Omega(\sigma,\Delta)$, we have*

$$\lim_{w\to\infty} p_{\Phi(w),\Delta}(\omega) = \begin{cases} 0 & \text{if } \omega \not\models \widehat{\Phi},\\ \frac{1}{|\{\omega\in\Omega(\sigma,\Delta) | \omega\models\widehat{\Phi}\}|} & \text{if } \omega \models \widehat{\Phi}. \end{cases}$$

Proof. Let $\tau(\omega) = \sum_{(\alpha_i,w)\in\Phi} w \cdot Q_\omega(\alpha_i)$. We have, for all worlds $\omega', \omega'' \in \Omega(\sigma,\Delta)$: if $Q_{\omega'}(\alpha) = 1$ for all $(\alpha, w) \in \Phi$ and $Q_{\omega''}(\beta) < 1$ for some $(\beta, w) \in \Phi$, then there is a positive real number ϵ such that

$$\frac{\tau(\omega') - \tau(\omega'')}{w} = \sum_{(\alpha_i,w)\in\Phi} Q_{\omega'}(\alpha_i) - \sum_{(\alpha_i,w)\in\Phi} Q_{\omega''}(\alpha_i) \geq \epsilon$$

It is easy to notice that $\tau(\omega') - \tau(\omega'') \to \infty$ for $w \to \infty$. Hence also

$$\frac{p_{\Phi(w),\Delta}(\omega')}{p_{\Phi(w),\Delta}(\omega'')} \to \infty, \text{ for } w \to \infty$$

Since, by Remark 1 above, for all α, $Q_\omega(\alpha) = 1$ iff $\omega \models \alpha'$ where α' is obtained by replacing all $\forall^*$ quantifiers in α by $\forall$, the proposition holds. □

4.3 Relational Marginal Problems and QMLNs

From the discussion in Section 2.3 and Point (5) in Definition 3, it follows that standard MLNs containing only quantifier-free FOL formulas are solutions of maximum entropy relational marginal problems constrained by sentence statistics of sentences that contain *only* $\forall^*$ quantifiers. Hence, a natural question is whether the same also holds for

relational marginal problems constrained by the more general sentence statistics given by Definition 3 which do contain statistical quantifiers $\forall^*$ as well as classical quantifiers $\forall$ and $\exists$. In the rest of the section we sketch the argument showing that the answer to this question is positive.

We use (α, q) to denote the constraint $Q[\alpha] = q$, where α is a quantified sentence and q is a probability. Let Δ be a finite domain and $\mathcal{C} = \{(\alpha_1, q_1), \ldots, (\alpha_k, q_k)\}$ be a set of constraints that the sought distribution must satisfy. We require some auxiliary notation. We define the sets

$$\mathcal{C}_H = \{\alpha \mid (\alpha, 1) \in \mathcal{C}\} \cup \{\mathrm{neg}\,(\alpha) \mid (\alpha, 0) \in \mathcal{C}\},$$
$$\mathcal{C}_S = \{(\alpha, w) \mid (\alpha, w) \in \mathcal{C} \text{ and } 0 < w < 1\}.$$

We further define a set Ω of worlds by taking

$$\Omega = \{\omega \in \Omega(\sigma, \Delta) \mid \bigwedge_{\alpha \in \mathcal{C}_H} Q_\omega(\alpha) = 1\}.$$

The motivation behind this is that, in any probability distribution $p(\omega)$ satisfying all constraints $\mathcal{C}$, every possible world *not in* Ω must have probability 0. We have to treat this separately from the rest of the constraints.[3]

In order to solve the relational marginal problem, we introduce a variable P_ω for every world $\omega \in \Omega$, intuitively representing the probability of ω. The optimization problem representing the maximum entropy relational marginal problem is then given by the objective function

$$\max_{\{P_\omega \mid \omega \in \Omega\}} \sum_{\omega \in \Omega} P_\omega \log \frac{1}{P_\omega}$$

subject to the following constraints:

$$\forall (\alpha_i, q_i) \in \mathcal{C}_S : \sum_{\omega \in \Omega} Q_\omega(\alpha_i) \cdot P_\omega = q_i$$
$$\forall \omega \in \Omega : P_\omega \geq 0, \quad \sum_{\omega \in \Omega} P_\omega = 1$$

Assuming[4] that there exists a feasible solution of the optimization problem such that $P_\omega > 0$ for all $\omega \in \Omega$, and using standard techniques from convex optimization (Boyd and Vandenberghe 2004; Singh and Vishnoi 2014), specifically the construction of Lagrangian dual problems and the use of Slater's condition, we arrive at the solution

$$P_\omega = \frac{1}{Z} \exp\left(\sum_{(\alpha_i, w_i) \in \mathcal{C}_S} w_i \cdot Q_\omega(\alpha_i)\right) \tag{7}$$

where Z is a normalization constant and the weights w_i are obtained as solutions of the optimization problem (the dual of the maximum entropy relational marginal problem) which is to maximize the following expression:

$$\sum_{(\alpha_i, q_i) \in \mathcal{C}_S} w_i q_i - \log\left(\sum_{\omega \in \Omega} \exp\left(\sum_{(\alpha_i, q_i) \in \mathcal{C}_S} w_i \cdot Q_\omega(\alpha_i)\right)\right)$$

[3] If we just plugged the constraints from $\mathcal{C}_H$ into the optimization problem, Slater's condition (Boyd and Vandenberghe 2004) would not hold and we could not establish the duality based on it.

[4] If this condition is not satisfied we have to add additional hard formulas that explicitly rule out the worlds that have zero probability in every solution satisfying the given marginal constraints.

As the main result of this section, we have verified that the maximum entropy relational entropy problem for QMLNs has a similar significance as in classical MLNs. In particular, both Points (i) and (ii) from Section 2.3 are satisfied as well for QMLNs, that is, QMLNs are also solutions of max entropy relational problems constrained by sentence statistics and the result of the latter agrees with maximum likelihood estimation when the input probabilities stem from a domain of the same size.

5 A Translation for MAP-Inference

In this section we describe a translation from *arbitrary* quantified MLNs to quantified MLNs that contain the statistical quantifiers $\forall^*$ only as a leading prefix. We have already seen in Proposition 1 that the latter QMLNs correspond essentially to standard MLNs. Since the translation can be performed in polynomial time, the given translation establishes a polynomial time reduction of MAP in QMLNs to MAP in MLNs.

Overview. The given quantified MLN is translated by processing the weighted sentences one by one. More specifically, we show how to eliminate a single classical quantifier that appears before a block of $\forall^*$'s in the quantifier prefix of the quantified sentence. By exhaustively applying this elimination, we end up with a set of weighted sentences where all $\forall^*$ quantifiers appear in a prefix block of $\forall^*$'s.

For the description of the elimination, let us suppose that (α, w) is a weighted sentence with α defined as follows

$$Q_1 x_1 \ldots Q_k x_k \forall^* x_{k+1} \ldots \forall^* x_{k+l} : \psi(x_1, \ldots, x_{k+l})$$

where $Q_i \in \{\exists, \forall, \forall^*\}$ for $1 \leq i \leq k-1$, $Q_k \in \{\exists, \forall\}$ and $\psi(x_1, \ldots, x_{k+l})$ is a formula with free variables $x_1, \ldots, x_{k+l}$; recall that the formula $\psi(x_1, \ldots, x_{k+l})$ may also contain variables bound by quantifiers $\exists$ and $\forall$ but not by $\forall^*$. The quantified sentence (α, w) is transformed into a set of hard constraints, that is, weighted sentences of the shape (φ, ∞) with φ a FOL sentence, and a single weighted sentence (α', w') with $w' = |\Delta| \cdot w$ and α' being

$$Q_1 x_1 \ldots Q_{k-1} x_{k-1} \forall^* x_k \ldots \forall^* x_{k+l} : \psi'(x_1, \ldots, x_{k+l})$$

for some formula ψ' to be defined below. Observe that the effect of the step is to turn quantifier Q_k into $\forall^*$.

Eliminating $Q_k x_k$ In order to simplify notation in the description of the elimination step, we will abbreviate $(x_1, \ldots, x_{k-1})$ with $\mathbf{x}$ and $(x_{k+1}, \ldots, x_{k+l})$ with $\mathbf{z}$, and write, e.g., $\psi(\mathbf{a}, a, \mathbf{b})$ instead of $\psi[\mathbf{x}/\mathbf{a}, x_k/a, \mathbf{z}/\mathbf{b}]$. We describe how to replace $Q_k x_k$ by $\forall^* x_k$. By Proposition 4, we can assume without loss of generality that Q_k is in fact $\exists$. By the semantics, the variable x_k *maximizes* the sentence statistics for the variables $\mathbf{z}$ over all possible choices of $a \in \Delta$. Our main idea is to simulate the computation of the sentence statistic in the MLN itself. For this purpose, we introduce a

fresh[5] k-ary predicate name *max*, set

$$\psi'(\mathbf{x}, x_k, \mathbf{z}) = \mathit{max}(\mathbf{x}, x_k) \wedge \psi(\mathbf{x}, x_k, \mathbf{z}),$$

and appropriately define *max* using hard constraints. More formally, let us denote with $\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a)$ the set of all assignments of $\mathbf{z}$ to values $\mathbf{b}$ such that $\psi(\mathbf{a}, a, \mathbf{b})$ is satisfied in world ω, that is,

$$\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a) = \{\mathbf{b} \in \Delta^l \mid \omega \models \psi(\mathbf{a}, a, \mathbf{b})\}.$$

Our goal is to enforce that, in every world ω, *max* satisfies the following property $(*)$:

$(*)$ for every choice $\mathbf{a} = (a_1, \ldots, a_{k-1})$ of values for $\mathbf{x}$, there is precisely one a^* such that $\omega \models \mathit{max}(\mathbf{a}, a^*)$, and moreover, this a^* satisfies

$$|\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a)| \leq |\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a^*)| \tag{8}$$

for all $a \in \Delta$.

Indeed, property $(*)$ formalizes the mentioned semantics for the sentence statistic for $\exists$. For enforcing it, observe that the inequality (8) is satisfied iff there is an injective mapping from the set on the left-hand side, $\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a)$, to the set on the right-hand side, $\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a^*)$. We exploit this observation as follows. First define a collection of linear orders on domain elements, one linear order $\preceq_{\mathbf{a}}$ for each assignment of a tuple $\mathbf{a}$ of domain elements to the variables in $\mathbf{x}$. We represent the order $\preceq_{\mathbf{a}}$ by the predicates $\mathit{leq}(\mathbf{a}, \cdot, \cdot)$. The linear orders are enforced by hard constraints. More precisely, we ensure that one such linear order exists for any assignment of domain elements to variables in $\mathbf{x}$ by adding hard constraints for axiomatizing antisymmetry, transitivity, and totality, respectively:

$$\forall \mathbf{x} \forall y, z \colon \mathit{leq}(\mathbf{x}, y, z) \wedge \mathit{leq}(\mathbf{x}, z, y) \Rightarrow y = z, \tag{9}$$

$$\forall \mathbf{x} \forall x, y, z \colon \mathit{leq}(\mathbf{x}, x, y) \wedge \mathit{leq}(\mathbf{x}, y, z) \Rightarrow \mathit{leq}(\mathbf{x}, x, z), \tag{10}$$

$$\forall \mathbf{x} \forall x, y \colon \mathit{leq}(\mathbf{x}, x, y) \vee \mathit{leq}(\mathbf{x}, y, x). \tag{11}$$

Next, we connect the linear order construction with the idea of injective mappings described above. This is done via another fresh predicate name *fn* which encodes the required mapping. Intuitively, in $\mathit{fn}(\mathbf{a}, a, a', \mathbf{b}, \mathbf{b}')$, $\mathbf{a}$ refers to the current assignment to $\mathbf{x}$, constants a, a' refer to the elements we are interested in for x_k, and the function maps $\mathbf{b} \in \mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a)$ to $\mathbf{b}' \in \mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a')$. We add the following hard constraints:

$$\forall \mathbf{x} \forall y, y' \forall \mathbf{z} \colon \mathit{leq}(\mathbf{x}, y, y') \wedge \psi(\mathbf{x}, y, \mathbf{z}) \Rightarrow (\exists \mathbf{z}' \colon \psi(\mathbf{x}, y', \mathbf{z}') \wedge \mathit{fn}(\mathbf{x}, y, y', \mathbf{z}, \mathbf{z}')),$$

$$\forall \mathbf{x} \forall y, y', z, z', z'' \colon \mathit{fn}(\mathbf{x}, y, y', \mathbf{z}, \mathbf{z}') \wedge \mathit{fn}(\mathbf{x}, y, y', \mathbf{z}, \mathbf{z}'') \Rightarrow \mathbf{z}' = \mathbf{z}'',$$

$$\forall \mathbf{x} \forall y, y', z, z', z'' \colon \mathit{fn}(\mathbf{x}, y, y', \mathbf{z}, \mathbf{z}') \wedge \mathit{fn}(\mathbf{x}, y, y', \mathbf{z}'', \mathbf{z}') \Rightarrow \mathbf{z} = \mathbf{z}''.$$

[5] In general, we need to introduce fresh predicates names for every transformed (α_i, w_i) separately, e.g. max_{α_i} etc. For brevity, we do not show this explicitly in the text.

The first two sentences enforce that, if $a \preceq_{\mathbf{a}} a'$, then there exists a mapping from $\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a)$ to $\mathsf{Wit}_{\psi,\omega}(\mathbf{a}, a')$. Injectivity of the mapping is ensured by the third sentence.

In order to define the predicate *max*, we add the following hard constraints:

$$\forall \mathbf{x} \exists y \colon \mathit{max}(\mathbf{x}, y),$$
$$\forall \mathbf{x}, y, y' \colon \mathit{max}(\mathbf{x}, y) \wedge \mathit{leq}(\mathbf{x}, y, y') \Rightarrow y = y'.$$

Correctness We have given some intuition above, but let us provide some more details. First, it is not hard to see that the added hard constraints ensure that *max* indeed satisfies the desired property $(*)$.

Now, let Δ be an arbitrary domain, ω^* be the most probable world of the QMLN Φ over domain Δ, and let Φ' be obtained from Φ by the application of a single quantifier elimination step. Further, denote with $\sigma' \supseteq \sigma$ the extended vocabulary. We call a world $\omega' \in \Omega(\sigma', \Delta)$ an *extension of* $\omega \in \Omega(\sigma, \Delta)$ if for every $R \in \sigma$ of arity k and all $a_1, \ldots, a_k \in \Delta$, we have $\omega' \models R(a_1, \ldots, a_k)$ iff $\omega \models R(a_1, \ldots, a_k)$. It is easy to see that our construction ensures that, in fact, every world in $\Omega(\sigma', \Delta)$ is the extension of a (unique!) world in $\Omega(\sigma, \Delta)$, and conversely, every world in $\Omega(\sigma, \Delta)$ has an extension in $\Omega(\sigma', \Delta)$. Moreover, the sentence statistics for α and its replacement α' relate as follows:

Lemma 6. *Let ω' be an extension of ω. Then*

$$Q_{\omega'}(\alpha') = \frac{1}{|\Delta|} Q_\omega(\alpha).$$

Proof. To see this, let $\mathbf{a}$ be some assignment to $\mathbf{x}$ and let $a^* \in \Delta$ be the element that exists by Property $(*)$. We have $Q_{\omega'}(\forall^* \mathbf{z} \colon \psi'(\mathbf{a}, a^*, \mathbf{z})) = Q_\omega(\forall^* \mathbf{z} \colon \psi(\mathbf{a}, a^*, \mathbf{z}))$ and, for all $a \neq a^*$, $Q_{\omega'}(\forall^* \mathbf{z} \colon \psi'(\mathbf{a}, a, \mathbf{z})) = 0$ since $\omega' \not\models \mathit{max}(\mathbf{a}, a)$ for such a. By the semantics, we obtain

$$Q_{\omega'}(\forall^* x_k \mathbf{z} \colon \psi'(\mathbf{a}, x_k, \mathbf{z})) = \frac{1}{|\Delta|} Q_\omega(\exists x_k \forall^* \mathbf{z} \colon \psi(\mathbf{a}, x_k, \mathbf{z}))$$

The statement from the Lemma follows from the fact that the constant factor $1/|\Delta|$ distributes over $\min, \max$ and expectation in the definition of the semantics of $\forall, \exists$, and $\forall^*$. □

The definition of the updated weight $w' = |\Delta| \cdot w$ now implies that there is a constant c such that $p_{\Phi',\Delta}(\omega') = c \cdot p_{\Phi,\Delta}(\omega)$ for all $\omega \in \Omega(\sigma, \Delta)$ and all extensions $\omega' \in \Omega(\sigma', \Delta)$ thereof. This establishes the correctness of the reduction.

Theorem 7. *If ω' is an extension of ω, then ω' is a most probable world in $p_{\Phi',\Delta}$ iff ω is a most probable world in $p_{\Phi,\Delta}$.*

6 A Translation for Marginal Inference

The translation given in the previous section does not quite work for marginal inference. Note that for MAP inference, it is enough if all extensions of any possible world of the original problem have the same weight. It is not a problem if there are multiple extensions of the same world or if different worlds have different numbers of extensions (as long

as they have the same weight). A single world sometimes has multiple extensions because the linear order *leq* and the functions represented by *fn* are not uniquely defined. However, for marginal inference, this is no longer acceptable. We have to ensure that any two worlds will have the same number of extensions.

We fix these problems by further restricting the functions encoded by *fn* and the order encoded by *leq*. More specifically, our goal is to add another set of hard constraints such that

$(**)$ every world $\omega \in \Omega(\sigma, \Delta)$ has the same number of extensions $\omega' \in \Omega(\sigma', \Delta)$.

To realize that, we exploit again the idea of the linear order. More specifically, we add a fresh binary predicate name $\leq$ and enforce that it is a linear order on Δ by using hard constraints such as those in Equations (9)–(11). Based on $\leq$, we break all possible ties that might occur in the definition of *max*, *leq*, *fn*, in the sense that for a fixed choice of $\leq$, there is exactly one choice of *max*, *leq*, *fn*. First, we enforce that *fn* has the right domain:

$$\forall \mathbf{x} \forall x, y \forall \mathbf{z}, \mathbf{z}' : \mathit{fn}(\mathbf{x}, x, y, \mathbf{z}, \mathbf{z}') \Rightarrow (\psi(\mathbf{x}, x, \mathbf{z}) \wedge \psi(\mathbf{x}, y, \mathbf{z}'))$$

For breaking ties in *leq*, we add the following constraint stating that, if $\mathit{leq}(\mathbf{x}, x, y)$ holds and the function encoded by *fn* is also surjective at given points $\mathbf{x}, x, y$, then $x \leq y$:

$$\forall \mathbf{x} \forall x, y : \Big(\mathit{leq}(\mathbf{x}, x, y) \wedge \big(\forall \mathbf{z}'.\psi(\mathbf{x}, y, \mathbf{z}') \Rightarrow \exists \mathbf{z}.\mathit{fn}(\mathbf{x}, x, y, \mathbf{z}, \mathbf{z}')\big)\Big) \Rightarrow x \leq y$$

Next we enforce that *fn* preserves the order $\leq$ by including the constraint

$$\forall \mathbf{x} \forall x, y \forall \mathbf{z}_1, \mathbf{z}_1', \mathbf{z}_2, \mathbf{z}_2' : \Big(\mathbf{z}_1 \leq^* \mathbf{z}_2 \wedge \mathit{fn}(\mathbf{x}, x, y, \mathbf{z}_1, \mathbf{z}_1') \wedge \mathit{fn}(\mathbf{x}, x, y, \mathbf{z}_2, \mathbf{z}_2')\Big) \Rightarrow \mathbf{z}_1' \leq^* \mathbf{z}_2'$$

where the order $\leq^*$ is defined – using straightforward constraints – as the (unique) lexicographic extension of $\leq$ to the arity of $\mathbf{z}$. For instance, consider the sets $\{1, 2\}$ and $\{1, 2, 3\}$. Assuming $1 \leq 2 \leq 3$, this constraint excludes, among others, the function that maps 1 to 3 and 2 to 1, because $\leq$ is not preserved: $1 \leq 2$, but $3 \not\leq 1$.

Finally, note that it can still be the case that two worlds have different number of extensions because each of the functions represented by *fn* is order-preserving w.r.t. $\leq^*$ and has a uniquely defined domain but, apart form that, does not have to satisfy any other constraints. For instance the number of such functions from $\{1, 2, 3\}$ to $\{1, 2, 3, 4\}$ and the number of such functions from $\{1, 2, 3\}$ to $\{1, 2, 3, 4, 5\}$ are different. We address this by requiring that the functions represented by *fn* map every element to the smallest element possible:

$$\forall \mathbf{x} \forall x, y \forall \mathbf{z}_1, \mathbf{z}_1' : \mathit{fn}(\mathbf{x}, x, y, \mathbf{z}_1, \mathbf{z}_1') \Rightarrow (\forall \mathbf{z}_2'.\psi(\mathbf{x}, y, \mathbf{z}_2') \wedge \mathbf{z}_2' \leq^* \mathbf{z}_1' \Rightarrow \exists \mathbf{z}_2.\mathit{fn}(\mathbf{x}, x, y, \mathbf{z}_2, \mathbf{z}_2'))$$

Correctness Let Φ'' be the result of adding the described constraints to Φ'. Based on the given intuitions, one can easily show that Property $(**)$ is satisfied. Since there are $|\Delta|!$ possible choices for $\leq$, we get that

$$p_{\Phi'',\Delta}(\omega') = \frac{1}{|\Delta|!} p_{\Phi,\Delta}(\omega),$$

for any extension ω' of ω. Since $\omega' \models \varphi$ iff $\omega \models \varphi$ for any given sentence φ over σ, we obtain the desired result:

Theorem 8. *For every FOL sentence φ, we have*

$$P_{\omega \sim p_{\Phi,\Delta}}(\varphi) = P_{\omega \sim p_{\Phi'',\Delta}}(\varphi).$$

An important question is whether the result from Theorem 8 can be extended to computing a marginal query $Q_{\Phi,\Delta}[\alpha]$ for a quantified sentence α. The answer to this question is positive. We next outline how this is done. Let Φ be the MLN resulting from applying the translation above to a given QMLN. We distinguish two cases.

(i) If α contains only $\forall^*$ quantifiers as a leading prefix of the block of quantifiers, we can borrow techniques from (Van Haaren et al. 2016). In particular, we need to create the partition of the groundings $\alpha\vartheta$ of α; specifically, groundings of the variables bound by $\forall^*$, such that the probability of any two groundings $\alpha\vartheta$ and $\alpha\vartheta'$ in the same equivalence class of this partition is equal, i.e. $P_{\omega \sim p_{\Phi,\Delta}}(\alpha\vartheta) = P_{\omega \sim p_{\Phi'',\Delta}}(\alpha\vartheta')$. This partitioning can be achieved using preemptive shattering (Poole, Bacchus, and Kisynski 2011). Once we have the partitioning, we can compute the marginal probability $P_{\omega \sim p_{\Phi,\Delta}}(\alpha\vartheta)$ for one representative $\alpha\vartheta_i$ from each partition class $\mathcal{P}_i$. Finally, we can compute $Q_{\Phi,\Delta}[\alpha]$ as follows:

$$Q_{\Phi,\Delta}[\alpha] = \frac{1}{|\bigcup_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i|} \sum_{\mathcal{P} \in \mathcal{P}_i} P_{\omega \sim p_{\Phi,\Delta}}(\alpha\vartheta_i) \cdot |\mathcal{P}_i|$$

(ii) If α contains an arbitrary quantifier prefix, we first need to convert it to the form assumed above, that is, we apply the transformation described in this section to obtain a sentence α' that contains $\forall^*$ quantifiers only as a prefix while also generating several additional hard constraints, which we include in Φ. We can then use the procedure outlined in (i) above to compute $Q_{\Phi,\Delta}[\alpha]$.

7 On QMLNs Restricted to Two Variables

It has been shown that marginal inference for MLNs can often be reduced to *symmetric weighted first-order model counting (WFOMC)*, see e.g. (Beame et al. 2015). In this context of particular importance is the *two-variable fragment of FOL (FO^2)*, since for FO^2 symmetric WFOMC can be solved in polynomial time data complexity, that is, when the formula is considered fixed and the only input is the domain Δ (Van den Broeck 2011; Van den Broeck, Meert, and Darwiche 2014; Beame et al. 2015). Given the fact that MLNs containing formulae with up two variables (2-MLNs) can be encoded as WFOMC for FO^2, 2-MLNs are *domain liftable*. Hence, a natural question to ask is whether the same holds for *quantified* MLNs.

Let us first remark that the reduction described in Section 6 does not preserve the quantifier rank. Indeed, consider

the example formula $\exists x \forall^* y \colon \mathsf{knows}(x, y)$. It is originally an FO^2 formula, but the elimination step introduces a *quaternary* predicate name *fn* and the constraints for this predicate require using *four* variables. As a consequence, we cannot 'reuse' the results on WFOMC for FO^2 and thus attain domain liftability for 2-QMLNs. Moreover, our reduction explicitly introduces transitivity axioms while there are only a few known very restricted cases where WFOMC is domain liftable in the presence of transitivity (Kazemi et al. 2016).

In the light of recent results by Kuusisto and Lutz (2018) on WFOMC for an extension of FO^2 with *counting quantifiers*, it is not very surprising that a straightforward translation *preserving* the quantifier rank from QMLNs to MLNs seems not possible. To see this, note that to compute the sentence statistic, we need to take into account the *out-degree* of domain elements, where the out-degree of an element is the number of elements that are related to it, cf. Example 1. Indeed, (Kuusisto and Lutz 2018) put quite some technical effort, relying on sophisticated model-theoretic techniques and combinatorics, to show that WFOMC for FO^2 with *one functional axiom*[6] is polynomial time in data complexity. The complexity of WFOMC for FO^2 with many functional axioms or more generally, with arbitrary counting quantifiers remains a challenging open problem. This gives an insight on the difficulty of studying the computational complexity of reasoning in QMLNs restricted to two variables, making it a research challenge by itself, which we plan to address in the future. In particular, we will investigate the exact connection between 2-QMLNs and WFOMC for extensions of FO^2 with some means for counting.

8 Related Work

Classical FOL quantifiers $(\forall, \exists)$ were already considered in the original work on Markov Logic Networks (Richardson and Domingos 2006), albeit without a rigorous definition. A precise treatment of FOL quantifiers was carried out later on in (Van den Broeck, Meert, and Darwiche 2014; Van den Broeck and Suciu 2017). In particular, Van den Broeck, Meert, and Darwiche (2014) show how to remove existential quantifiers while preserving marginal inference results. In all these works, however, MLNs with quantifiers were defined in a way that is equivalent to QMLNs with a prefix of $\forall^*$ quantifiers. As a consequence, it is not possible to directly represent statistics (*features* in MLN parlance) that correspond to sentences in which $\forall$ or $\exists$ precedes $\forall^*$ in the quantifier block.

There has been also some work on other types of aggregation. For example, some works considered explicit constructs for counting in relational models (Milch et al. 2008; Jain, Barthels, and Beetz 2010). In another research effort, Lowd and Domingos (2007) introduced recursive random fields that are capable of emulating certain forms of more complex aggregation. However, recursive random fields do not seem capable of even representing statistics such as $\exists x \forall^* y : \mathsf{knows}(x, y)$. Finally, Beltagy and Erk (2015) studied the effect of the domain closure assumption on the semantics of probabilistic logic when existential quantifiers are allowed.

In the context of *probabilistic soft logic (PSL)*, Farnadi et al. (2017) recently introduced soft quantifiers based on quantifiers from *fuzzy logic*. However, their approach strictly applies to fuzzy logic. In particular, in PSL random variables e.g. $\mathsf{smokes}(\mathsf{Alice})$, may acquire non-Boolean truth values.

Another stream of research that is related to our work is the study of the effect of domain size and its extrapolation on the probability distributions encoded using various relational learning systems (Poole et al. 2014; Kazemi et al. 2014; Kuželka et al. 2018). However, none of these works studied the interplay of statistical and classical quantifiers.

There has been also some related work within KR research, e.g. about statistical reasoning in description logics (Peñaloza and Potyka 2017; Lutz and Schröder 2010).

9 Discussion and Future Work

In this paper, we have investigated the extension QMLNs of MLNs with statistical quantifiers, allowing to express e.g. measures on the proportion of domain elements fulfilling certain property. We developed some key foundations by establishing a relation between MLNs and QMLNs. In particular, we provided a polytime reduction of the standard reasoning tasks MAP and MARG in QMLNs to their counterpart in MLNs. Furthermore, we also showed how to generalize the random substitution semantics to QMLNs.

As for future work, it might be interesting to develop more direct approaches to MAP and MARG in QMLNs. Indeed, even if the developed translations provide polytime reductions of reasoning in QMLNs to reasoning in MLNs (and overall, a good understanding of the relation between QMLNs and MLNs), they do not yield an immediate practical approach since the introduction of new predicates with greater arity is required. Another interesting aspect of future work is to investigate the statistical properties of QMLNs. For MLNs with quantifier-free formulas, Kuželka et al. (2018) derived bounds on expected errors of the statistics' estimates. However, obtaining similar bounds for the more general statistics considered here seems considerably more difficult because of the minimization and maximization that are involved in them.

Acknowledgments

The authors were supported by EU's Horizon 2020 programme under the Marie Skłodowska-Curie grant 663830, ERC consolidator grant 647289 CODA, and the Research Foundation - Flanders (project G.0428.15), respectively.

References

Bacchus, F.; Grove, A. J.; Koller, D.; and Halpern, J. Y. 1992. From statistics to beliefs. In *Proc. of AAAI-92*, 602–608.

Beame, P.; Van den Broeck, G.; Gribkoff, E.; and Suciu, D. 2015. Symmetric weighted first-order model counting. In *Proc. of PODS-15*, 313–328.

[6] Intuitively, enforcing out-degree 1 for a particular binary relation, that is, elements are related with at most one element.

Beltagy, I., and Erk, K. 2015. On the proper treatment of quantifiers in probabilistic logic semantics. In *Proc. of IWCS-15*, 140–150.

Boyd, S., and Vandenberghe, L. 2004. *Convex optimization*. Cambridge university press.

Chen, C.; Kolmogorov, V.; Zhu, Y.; Metaxas, D. N.; and Lampert, C. H. 2013. Computing the M most probable modes of a graphical model. In *Proc. of AISTATS-13*, 161–169.

Farasat, A.; Nikolaev, A. G.; Srihari, S. N.; and Blair, R. H. 2015. Probabilistic graphical models in modern social network analysis. *Social Netw. Analys. Mining* 5(1):62:1–62:18.

Farnadi, G.; Bach, S. H.; Moens, M.; Getoor, L.; and Cock, M. D. 2017. Soft quantification in statistical relational learning. *Machine Learning* 106(12):1971–1991.

Jain, D.; Barthels, A.; and Beetz, M. 2010. Adaptive markov logic networks: Learning statistical relational models with dynamic parameters. In *Proc. of ECAI 2010*, 937–942.

Kazemi, S. M.; Buchman, D.; Kersting, K.; Natarajan, S.; and Poole, D. 2014. Relational logistic regression. In *Proc. of KR-14*.

Kazemi, S. M.; Kimmig, A.; den Broeck, G. V.; and Poole, D. 2016. New liftable classes for first-order probabilistic inference. In *Proc. of NIPS-16*, 3117–3125.

Kuusisto, A., and Lutz, C. 2018. Weighted model counting beyond two-variable logic. In *Proc. of LICS-18*, 619–628.

Kuželka, O.; Wang, Y.; Davis, J.; and Schockaert, S. 2018. Relational marginal problems: Theory and estimation. In *Proc. of AAAI-18*.

Lowd, D., and Domingos, P. M. 2007. Recursive random fields. In *Proc. of IJCAI-07*, 950–955.

Lutz, C., and Schröder, L. 2010. Probabilistic description logics for subjective uncertainty. In *Proc. of KR-10*.

Milch, B.; Zettlemoyer, L. S.; Kersting, K.; Haimes, M.; and Kaelbling, L. P. 2008. Lifted probabilistic inference with counting formulas. In *Proc. of AAAI-08*, 1062–1068.

Peñaloza, R., and Potyka, N. 2017. Towards statistical reasoning in description logics over finite domains. In *Proc. of SUM-17*, 280–294.

Poole, D.; Bacchus, F.; and Kisynski, J. 2011. Towards completely lifted search-based probabilistic inference. *arXiv preprint arXiv:1107.4035*.

Poole, D.; Buchman, D.; Kazemi, S. M.; Kersting, K.; and Natarajan, S. 2014. Population size extrapolation in relational probabilistic modelling. In *Proc. of SUM-14*, 292–305.

Richardson, M., and Domingos, P. M. 2006. Markov logic networks. *Machine Learning* 62(1-2):107–136.

Riedel, S., and McCallum, A. 2011. Robust biomedical event extraction with dual decomposition and minimal domain adaptation. In *Proc. of BioNLP-11*, 46–50.

Schulte, O.; Khosravi, H.; Kirkpatrick, A. E.; Gao, T.; and Zhu, Y. 2014. Modelling relational statistics with Bayes nets. *Machine Learning* 94(1):105–125.

Shalizi, C. R., and Rinaldo, A. 2013. Consistency under sampling of exponential random graph models. *The Annals of Statistics* 41(2):508–535.

Singh, M., and Vishnoi, N. K. 2014. Entropy, optimization and counting. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 50–59. ACM.

Tran, S. D., and Davis, L. S. 2008. Event modeling and recognition using markov logic networks. In *Proc. of ECCV-08*, 610–623.

Van den Broeck, G., and Suciu, D. 2017. *Query Processing on Probabilistic Data: A Survey*. Foundations and Trends in Databases. Now Publishers.

Van den Broeck, G.; Meert, W.; and Darwiche, A. 2014. Skolemization for weighted first-order model counting. In *Proc. of KR-14*.

Van den Broeck, G. 2011. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Proc. of NIPS-11*, 1386–1394.

Van Haaren, J.; Van den Broeck, G.; Meert, W.; and Davis, J. 2016. Lifted generative learning of Markov logic networks. *Machine Learning* 103(1):27–55.

Venugopal, D.; Chen, C.; Gogate, V.; and Ng, V. 2014. Relieving the computational bottleneck: Joint inference for event extraction with high-dimensional features. In *Proc. of EMNLP-14*, 831–843.