Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference*

[Highlight on Published Work]

Mathias Niepert
Computer Science and Engineering
University of Washington
Seattle, USA
mniepert@cs.washington.edu

Guy Van den Broeck†
Computer Science Department
University of California
Los Angeles, USA
guyvdb@cs.ucla.edu

ABSTRACT

Exchangeability is a central notion in statistics and probability theory. The assumption that an infinite sequence of data points is exchangeable is at the core of Bayesian statistics. However, *finite* exchangeability as a statistical property that renders probabilistic inference tractable is less well-understood. We develop a theory of finite exchangeability and its relation to tractable probabilistic inference. The theory is complementary to that of independence and conditional independence. We show that tractable inference in probabilistic models with high treewidth and millions of variables can be explained with the notion of finite (partial) exchangeability. We also show that existing lifted inference algorithms implicitly utilize a combination of conditional independence and partial exchangeability.

1. INTRODUCTION

Probabilistic graphical models such as Bayesian and Markov networks explicitly represent *conditional independencies* of a probability distribution with their structure [24; 16; 14; 4]. Their wide-spread use in research and industry can largely be attributed to this structural property and their declarative nature, separating representation and inference algorithms. Conditional independencies often lead to a more concise representation and facilitate efficient algorithms for parameter estimation and probabilistic inference. It is well-known, for instance, that probabilistic graphical models with a tree structure admit efficient inference. In addition to conditional independencies, modern inference algorithms exploit *contextual independencies* [1] to speed up probabilistic inference.

*The paper appears in the proceedings of the 28th Conference on Artificial Intelligence (AAAI), 2013.

†Both authors contributed equally to the paper. Guy Van den Broeck is also at KU Leuven, Belgium.

The time complexity of classical probabilistic inference algorithms is exponential in the treewidth [27] of the graphical model. Independence and its various manifestations often reduce treewidth and treewidth has been used in the literature as the decisive factor for assessing the tractability of probabilistic inference (cf. [14], [4]). However, recent algorithmic developments have shown that inference in probabilistic graphical models can be highly tractable, even in high-treewidth models without any conditional independencies. For instance, lifted probabilistic inference algorithms [25; 13] often perform efficient inference in densely connected graphical models with millions of random variables. With the success of lifted inference, understanding these algorithms and their tractability on a more fundamental level has become a central challenge. The most pressing question concerns the underlying statistical principle that allows inference to be tractable in the absence of independence.

The present paper contributes to a deeper understanding of the statistical properties that render inference tractable. We consider an inference problem tractable when it is solved by an efficient algorithm, running in time polynomial in the number of random variables. The crucial contribution is a comprehensive theory that relates the notion of finite partial *exchangeability* [7] to tractability. One instance is full exchangeability where the distribution is invariant under variable permutations. We develop a theory of exchangeable decompositions that results in novel tractability conditions. Similar to conditional independence, partial exchangeability decomposes a probabilistic model so as to facilitate efficient inference. Most importantly, the notions of conditional independence and partial exchangeability are complementary, and when combined, define a much larger class of tractable models than the class of models rendered tractable by conditional independence alone.

Conditional and contextual independence are such powerful concepts because they are *statistical* properties of the distribution, regardless of the representation used. Partial exchangeability is such a statistical property that is independent of any representation, be it a joint probability table, a Bayesian network, or a statistical relational model. We introduce novel forms of exchangeability, discuss their sufficient statistics, and efficient inference algorithms. The resulting exchangeability framework allows us to state known

liftability results as corollaries, providing a first statistical characterization of exact lifted inference. As an additional contribution, we connect the semantic notion of exchangeability to *syntactic* notions of tractability by showing that liftable statistical relational models have the required exchangeability properties due to their syntactic symmetries. We thereby unify notions of lifting from the exact and approximate inference community into a common framework.

2. A CASE STUDY: MARKOV LOGIC

The analysis of exchangeability and tractable inference is developed in the context of arbitrary discrete probability distributions, independent of a particular representational formalism. Nevertheless, for the sake of accessibility, we will provide examples and intuitions for Markov logic, a well-known statistical relational language that exhibits several forms of exchangeability. Hence, after the derivation of the theoretical results in each section, we apply the theory to the problem of inference in Markov logic networks. This also allows us to link the theory to existing results from the lifted probabilistic inference literature.

2.1 Markov Logic Networks

We first introduce some standard concepts from function-free first-order logic. An *atom* $P(t_1, \ldots, t_n)$ consists of a predicate P/n of arity n followed by n argument terms t_i, which are either *constants*, $\{\mathsf{A}, \mathsf{B}, \ldots\}$ or *logical variables* $\{\mathsf{x}, \mathsf{y}, \ldots\}$. A *unary atom* has one argument and a *binary atom* has two. A formula combines atoms with connectives (e.g., $\wedge$, $\Leftrightarrow$). A formula is *ground* if it contains no logical variables. The groundings of a formula are obtained by instantiating the variables with particular constants.

Many statistical relational languages have been proposed in recent years [9; 5]. We will work with one such language, called *Markov logic networks* (MLN) [26]. An MLN is a set of tuples (w, f), where w is a real number representing a weight and f is a formula in first-order logic. Let us, for example, consider the following MLN

$$1.3 \quad \texttt{Smokes}(\mathsf{x}) \Rightarrow \texttt{Cancer}(\mathsf{x}) \tag{1}$$

$$1.5 \quad \texttt{Smokes}(\mathsf{x}) \wedge \texttt{Friends}(\mathsf{x}, \mathsf{y}) \Rightarrow \texttt{Smokes}(\mathsf{y}) \tag{2}$$

which states that smokers are more likely to (1) get cancer and (2) be friends with other smokers.

Given a domain of constants $\mathbf{D}$, a first-order MLN Δ induces a *grounding*, which is the MLN obtained by replacing each formula in Δ with all its groundings (using the same weight). Take for example the domain $\mathbf{D} = \{\mathsf{A}, \mathsf{B}\}$ (e.g., two people, Alice and Bob), the above first-order MLN represents the following grounding.

$$1.3 \quad \texttt{Smokes}(\mathsf{A}) \Rightarrow \texttt{Cancer}(\mathsf{A})$$
$$1.3 \quad \texttt{Smokes}(\mathsf{B}) \Rightarrow \texttt{Cancer}(\mathsf{B})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{A}) \wedge \texttt{Friends}(\mathsf{A}, \mathsf{A}) \Rightarrow \texttt{Smokes}(\mathsf{A})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{A}) \wedge \texttt{Friends}(\mathsf{A}, \mathsf{B}) \Rightarrow \texttt{Smokes}(\mathsf{B})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{B}) \wedge \texttt{Friends}(\mathsf{B}, \mathsf{A}) \Rightarrow \texttt{Smokes}(\mathsf{A})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{B}) \wedge \texttt{Friends}(\mathsf{B}, \mathsf{B}) \Rightarrow \texttt{Smokes}(\mathsf{B})$$

This ground MLN contains eight different random variables, which correspond to all groundings of atoms $\texttt{Smokes}(\mathsf{x})$, $\texttt{Cancer}(\mathsf{x})$ and $\texttt{Friends}(\mathsf{x}, \mathsf{y})$. This leads to a distribution over 2^8 possible worlds. The weight of each world is the product of the expressions $\exp(w)$, where (w, f) is a ground MLN formula and f is satisfied by the world. The probabilities of worlds are obtained by normalizing their weights. Without loss of generality [12], we assume that first-order formulas contain no constants.

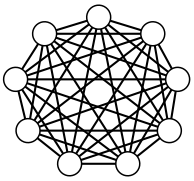


Figure 1: An undirected graphical model with 9 finitely exchangeable Bernoulli variables. There are no (conditional) independencies that hold among the variables.

2.2 Lifted Probabilistic Inference

The advent of statistical relational languages such as Markov logic has motivated a new class of *lifted inference* algorithms [25]. These algorithms exploit the high-level structure and symmetries of the first-order logic formulas to speed up inference [13]. Surprisingly, they perform tractable inference even in the absence of conditional independencies. For example, when interpreting the above MLN as an (undirected) probabilistic graphical model, all pairs of random variables in $\{\texttt{Smokes}(\mathsf{A}), \texttt{Smokes}(\mathsf{B}), \ldots\}$ are connected by an edge due to the groundings of Formula 2. The model has no conditional or contextual independencies between the `Smokes` variables. Nevertheless, lifted inference algorithms exactly compute its single marginal probabilities in time linear in the size of the corresponding graphical model [34], scaling up to millions of random variables.

As lifted inference research makes algorithmic progress, the quest for the source of tractability and its theoretical properties becomes increasingly important. For exact lifted inference, most theoretical results are based on the notion of domain-lifted inference [31].

Definition 1 (Domain-lifted). *Domain-lifted inference algorithms run in time polynomial in* $|\mathbf{D}|$.

Note that domain-lifted algorithms can be exponential in other parameters, such as the number of formulas and predicates. Our current understanding of exact lifted inference is that syntactic properties of MLN formulas permit domain-lifted inference [31; 11; 30]. We will review these results where relevant. Moreover, the (fractional) automorphisms of the graphical model representation have been related to lifted inference [21; 3; 23; 19]. While there are deep connections between automorphisms and exchangeability [21; 22; 3; 2], we refer these to future work.

3. FINITE EXCHANGEABILITY

This section provides some background on the concept of finite partial exchangeability. We proceed by showing that particular forms of finite exchangeability permit tractable

inference. For the sake of simplicity and to provide links to statistical relational models such as MLNs, we present the theory for *finite* sets of (upper-case) *binary* random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$. However, the theory applies to all distributions over finite valued discrete random variables. Lower-case $\mathbf{x}$ denote an assignments to $\mathbf{X}$.

We begin with the most basic form of exchangeability.

DEFINITION 2 (FULL EXCHANGEABILITY). *A set of variables* $\mathbf{X} = \{X_1, ..., X_n\}$ *is* fully exchangeable *if and only if* $\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_1 = x_{\pi(1)}, \ldots, X_n = x_{\pi(n)})$ *for all permutations* π *of* $\{1, \ldots, n\}$.

Full exchangeability is best understood in the context of a finite sequence of binary random variables such as a number of coin tosses. Here, exchangeability means that it is only the number of heads that matters and not their particular order. Figure 1 depicts an undirected graphical model with 9 finitely exchangeable dependent Bernoulli variables.

3.1 Finite Partial Exchangeability

The assumption that all variables of a probabilistic model are exchangeable is often too strong. Fortunately, exchangeability can be generalized to the concept of partial exchangeability using the notion of a *sufficient statistic* [8; 17; 15]. Particular instances of exchangeability such as full finite exchangeability correspond to particular statistics.

DEFINITION 3 (PARTIAL EXCHANGEABILITY). *Let* $\mathcal{D}_i$ *be the domain of* X_i*, and let* $\mathcal{T}$ *be a finite set. A set of random variables* $\mathbf{X}$ *is* partially exchangeable *with respect to the statistic* $T : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to \mathcal{T}$ *if and only if*

$$T(\mathbf{x}) = T(\mathbf{x}') \text{ implies } \Pr(\mathbf{x}) = \Pr(\mathbf{x}').$$

The following theorem states that the joint distribution of a set of random variables that is partially exchangeable with a statistic T is a unique mixture of uniform distributions.

THEOREM 1 (*[7]*). *Let* $\mathcal{T}$ *be a finite set and let* $T : \{0,1\}^n \to \mathcal{T}$ *be a statistic of a partially exchangeable set* $\mathbf{X}$*. Moreover, let* $S_t = \{\mathbf{x} \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \mid T(\mathbf{x}) = t\}$*, let* U_t *be the uniform distribution over* S_t*, and let* $\Pr(S_t) = \Pr(T(\mathbf{x}) = t)$*. Then,*

$$\Pr(\mathbf{X}) = \sum_{t \in \mathcal{T}} \Pr(S_t) U_t(\mathbf{X}).$$

Hence, a distribution that is partially exchangeable with respect to a statistic T can be parameterized as a unique mixture of uniform distributions. We will see that several instances of partial exchangeability render probabilistic inference tractable. Indeed, the major theme of the present paper can be summarized as finding methods for constructing the above representation and exploiting it for tractable probabilistic inference for a given probabilistic model.

Let $[[\cdot]]$ be the indicator function. The uniform distribution of each equivalence class S_t is $U_t(\mathbf{X}) = [[T(\mathbf{X}) = t]]/|S_t|$; and the probability of S_t is $\Pr(S_t) = \Pr(\mathbf{x})|S_t|$ for every $\mathbf{x} \in S_t$. Hence, every value of the statistic T corresponds to one equivalence class S_t of joint assignments with identical probability. We will refer to these equivalence classes as *orbits*. We write $\mathbf{x} \sim \mathbf{e}$ when assignments $\mathbf{x}$ and $\mathbf{e}$ agree on the values of their shared variables [4]. The *suborbit* $S_{t,\mathbf{e}} \subseteq S_t$ for some evidence state $\mathbf{e}$ is the set of those states in S_t that are compatible with $\mathbf{e}$, that is, $S_{t,\mathbf{e}} = \{\mathbf{x} \mid T(\mathbf{x}) = t \text{ and } \mathbf{x} \sim \mathbf{e}\}$.

3.2 Partial Exchangeability and Probabilistic Inference

We are now in the position to relate finite partial exchangeability to tractable probabilistic inference, using notions from Theorem 1. The inference tasks we consider are

- *MPE inference*, i.e., finding $\arg\max_{\mathbf{y}} \Pr(\mathbf{y}, \mathbf{e})$ for any given assignment $\mathbf{e}$ to variables $\mathbf{E} \subseteq \mathbf{X}$, and
- *marginal inference*, i.e., computing $\Pr(\mathbf{e})$ for any given $\mathbf{e}$.

For a set of variables $\mathbf{X}$, we say that $P(\mathbf{x})$ can be computed *efficiently* iff it can be computed in time polynomial in $|\mathbf{X}|$. We make the following complexity claims

THEOREM 2. *Let* $\mathbf{X}$ *be partially exchangeable with statistic* T*. If we can efficiently*

- *for all* $\mathbf{x}$*, evaluate* $\Pr(\mathbf{x})$*, and*
- *for all* $\mathbf{e}$ *and* $t \in \mathcal{T}$ *decide whether there exists an* $\mathbf{x} \in S_{t,\mathbf{e}}$*, and if so, construct it,*

then the complexity of MPE inference *is polynomial in* $|\mathcal{T}|$*. If we can additionally compute* $|S_{t,\mathbf{e}}|$ *efficiently, then the complexity of* marginal inference *is also polynomial in* $|\mathcal{T}|$*.*

PROOF. For MPE inference, we construct an $\mathbf{x}_t \in S_{t,\mathbf{e}}$ for each $t \in \mathcal{T}$, and return the one maximizing $\Pr(\mathbf{x}_t)$. For marginal inference, we return $\sum_{t \in \mathcal{T}} \Pr(\mathbf{x}_t)|S_{t,\mathbf{e}}|$. □

If the above conditions for tractable inference are fulfilled we say that a distribution is *tractably* partially exchangeable for MPE or marginal inference. We will present notions of exchangeability and related statistics T that make distributions tractably partially exchangeable. Please note that Theorem 2 generalizes to situations in which we can only efficiently compute $\Pr(\mathbf{x})$ up to a constant factor Z, as is often the case in undirected probabilistic graphical models.

3.3 Markov Logic Case Study

Exchangeability and independence are not mutually exclusive. Independent and identically distributed (iid) random variables are also exchangeable. Take for example the MLN

1.5 `Smokes(x)`

The random variables $\texttt{Smokes}(\mathsf{A}), \texttt{Smokes}(\mathsf{B}), \ldots$ are independent. Hence, we can compute their marginal probabilities independently as

$$\Pr(\texttt{Smokes}(\mathsf{A})) = \Pr(\texttt{Smokes}(\mathsf{B})) = \frac{\exp(1.5)}{\exp(1.5)+1}$$

The variables are also finitely exchangeable. For example, the probability that A smokes and B does not is equal to the probability that B smokes and A does not. The sufficient statistic $T(\mathbf{x})$ counts *how many* people smoke in the state $\mathbf{x}$ and the probability of a state in which n out of N people smoke is $\exp(1.5n)/(\exp(1.5)+1)^N$.

Exchangeability can occur without independence, as in the following Markov logic network

$$1.5 \quad \texttt{Smokes}(\mathsf{x}) \wedge \texttt{Smokes}(\mathsf{y})$$

This distribution has neither independencies nor conditional independencies. However, its variables are finitely exchangeable and the probability of a state $\mathbf{x}$ is only a function of the sufficient statistic $T(\mathbf{x})$ counting the number of smokers in $\mathbf{x}$. The probability of a state now increases by a factor of $\exp(1.5)$ with every *pair of smokers*. When n people smoke there are n^2 pairs and, hence, $\Pr(\mathbf{x}) = \exp\left(1.5n^2\right)/Z$, where Z is a normalization constant. Let $\mathbf{Y}$ consist of all $\texttt{Smokes}(\mathsf{x})$ variables except for $\texttt{Smokes}(\mathsf{A})$, and let $\mathbf{y}$ be an assignment to $\mathbf{Y}$ in which m people smoke. The probability that A smokes given $\mathbf{y}$ is

$$\Pr\left(\texttt{Smokes}(\mathsf{A}) \mid \mathbf{y}\right) = \frac{\exp\left(1.5(m+1)^2\right)}{\exp\left(1.5(m+1)^2\right) + \exp\left(1.5m^2\right)},$$

which clearly depends on the number of smokers in $\mathbf{y}$. Hence, $\texttt{Smokes}(\mathsf{A})$ is not independent of $\mathbf{Y}$ but the random variables are exchangeable with sufficient statistic n. Figure 1 depicts the graphical representation of the corresponding ground Markov logic network.

4. EXCHANGEABLE DECOMPOSITIONS

We now present novel instances of partial exchangeability that render probabilistic inference tractable. These instances generalize exchangeability of single variables to exchangeability of sets of variables. We describe the notion of an exchangeable decomposition and prove that it fulfills the tractability requirements of Theorem 2. We proceed by demonstrating that these forms commonly occur in MLNs.

4.1 Variable Decompositions

The notions of independent and exchangeable decompositions are at the core of the developed theoretical results.

Definition 4 (Variable Decomposition). *A* variable decomposition $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_k\}$ *partitions* $\mathbf{X}$ *into subsets* $\mathbf{X}_i$. *We call* $w = \max_i |\mathbf{X}_i|$ *the* width *of the decomposition.*

Definition 5 (Independent Decomposition). *A variable decomposition is* independent *if and only if* Pr *factorizes as*

$$\Pr(\mathbf{X}) = \prod_{i=1}^{k} Q_i(\mathbf{X}_i).$$

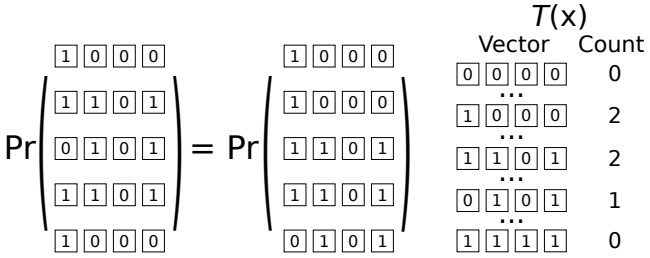


Figure 2: An exchangeable decomposition of 20 binary random variables (the boxes) into 5 components of width 4 (the rows). The statistic T counts the occurrences of each unique binary vector.

Definition 6 (Exchangeable Decomposition). *A variable decomposition is* exchangeable *iff for all permutations* π*,*

$$\Pr\left(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_k = \mathbf{x}_k\right) = \Pr\left(\mathbf{X}_1 = \mathbf{x}_{\pi(1)}, \ldots, \mathbf{X}_k = \mathbf{x}_{\pi(k)}\right).$$

Figure 2 depicts an example distribution Pr with 20 random variables and a decomposition into 5 subsets of width 4. The joint distribution is invariant under permutations of the 5 sequences. The corresponding sufficient statistic T counts the number of occurrences of each binary vector of length 4 and returns a tuple of counts.

Please note that the definition of full finite exchangeability (Definition 2) is the special case when the exchangeable decomposition has width 1. Also note that the size of all subsets in an exchangeable decomposition equal the width.

4.2 Tractable Variable Decompositions

The core observation of the present work is that variable decompositions that are exchangeable and/or independent result in tractable probabilistic models. For independent decompositions, the following tractability guarantee is used in most existing inference algorithms.

Proposition 3. *Given an independent decomposition of* $\mathbf{X}$ *with bounded width, and a corresponding factorized representation of the distribution (cf. Definition 5), the complexity of MPE and marginal inference is polynomial in* $|\mathbf{X}|$.

While the decomposition into independent components is a well-understood concept, the combination with finite exchangeability has not been previously investigated as a statistical property that facilitates tractable probabilistic inference. We can now prove the following result.

Theorem 4. *Suppose we can compute* $\Pr(\mathbf{x})$ *in time polynomial in* $|\mathbf{X}|$. *Then, given an exchangeable decomposition of* $\mathbf{X}$ *with bounded width, the complexity of MPE and marginal inference is polynomial in* $|\mathbf{X}|$.

Proof. Following Theorem 2, we have to show that there exists a statistic T so that (a) $|\mathcal{T}|$ is polynomial in $|\mathbf{X}|$; (b) we

can efficiently decide whether an $\mathbf{x} \in S_{t,\mathbf{e}}$ exists and if so, construct it; and (c) efficiently compute $|S_{t,\mathbf{e}}|$ for all $\mathbf{e}$ and $t \in \mathcal{T}$. Statements (b) and (c) ensure that the assumptions of Theorem 2 hold for exchangeable decompositions, which combined with (a) proves the theorem.

To prove (a), let us first construct a sufficient statistic for exchangeable decompositions. A full joint assignment $\mathbf{X} = \mathbf{x}$ decomposes into assignments $\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_k = \mathbf{x}_k$ in accordance with the given variable decomposition. Each $\mathbf{x}_\ell$ is a bit string $b \in \{0,1\}^w$. Consider a statistic $T(\mathbf{x}) = (c_1, \ldots, c_{2^w})$, where each c_i has a corresponding unique bit string $b_i \in \{0,1\}^w$ and $c_i = \sum_{\ell=1}^{k} [[\mathbf{x}_\ell = b_i]]$. The value c_i of the statistic thus represents the number of components in the decomposition that are assigned bit string b_i. Hence, we have $\sum_{i=1}^{2^w} c_i = k$, and we prove (a) by observing that

$$|\mathcal{T}| = \binom{k + 2^w - 1}{2^w - 1} \leq k^{2^w - 1} \leq n^{2^w - 1}.$$

To prove statements (b) and (c) we have to find, for each partial assignment $\mathbf{E} = \mathbf{e}$, an algorithm that generates an $\mathbf{x} \in S_{t,\mathbf{e}}$ and that computes $|S_{t,\mathbf{e}}|$ in time polynomial in $|\mathbf{X}|$. To hint at the proof strategy, we give the formula for the orbit without evidence $|S_t|$:

$$\left|S_{(c_1,\ldots,c_{2^w})}\right| = \binom{k}{c_1}\binom{k-c_1}{c_2}\cdots\binom{k-\sum_{i=1}^{2^w-2} c_i}{c_{2^w-1}}.$$

The proof is very technical and deferred to the appendix. □

4.3 Markov Logic Case Study

Let us consider the following MLN

$$1.3 \quad \texttt{Smokes}(\mathsf{x}) \Rightarrow \texttt{Cancer}(\mathsf{x})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{x}) \wedge \texttt{Smokes}(\mathsf{y})$$

It models a distribution in which every non-smoker or smoker with cancer, that is, every x satisfying the first formula, increases the probability by a factor of exp(1.3). Each pair of smokers increases the probability by a factor of exp(1.5). This model is not fully exchangeable: swapping $\texttt{Smokes}(\mathsf{A})$ and $\texttt{Cancer}(\mathsf{A})$ in a state yields a different probability. There are no (conditional) independencies between the $\texttt{Smokes}(\mathsf{x})$ atoms.

The variables in this MLN do have an exchangeable decomposition whose width is two, namely

$$\begin{aligned}\mathcal{X} = \{&\{\texttt{Smokes}(\mathsf{A}), \texttt{Cancer}(\mathsf{A})\},\\ &\{\texttt{Smokes}(\mathsf{B}), \texttt{Cancer}(\mathsf{B})\},\\ &\{\texttt{Smokes}(\mathsf{C}), \texttt{Cancer}(\mathsf{C})\}, \ldots\}.\end{aligned}$$

The sufficient statistic of this decomposition counts the number of people in each of four groups, depending on whether they smoke, and whether they have cancer. The probability of a state only depends on the number of people in each group and swapping people between groups does not change the probability of a state. For example,

$$\begin{aligned}&\Pr(\texttt{Smokes}(\mathsf{A}) = 0, \texttt{Cancer}(\mathsf{A}) = 1,\\ &\qquad\qquad \texttt{Smokes}(\mathsf{B}) = 1, \texttt{Cancer}(\mathsf{B}) = 0)\\ &= \Pr(\texttt{Smokes}(\mathsf{A}) = 1, \texttt{Cancer}(\mathsf{A}) = 0,\\ &\qquad\qquad \texttt{Smokes}(\mathsf{B}) = 0, \texttt{Cancer}(\mathsf{B}) = 1).\end{aligned}$$

Theorem 4 says that this MLN permits tractable inference.

The fact that this MLN has an exchangeable decomposition is not a coincidence. In general, we can show this for MLNs of unary atoms, which are called *monadic* MLNs.

THEOREM 5. *The variables in a monadic MLN have an exchangeable decomposition. The width of this decomposition is equal to the number of predicates.*

The proof builds on syntactic symmetries of the MLN, called renaming automorphisms [3; 20]. Please see the appendix for further details.

It now follows as a corollary of Theorems 4 and 5 that MPE and marginal inference in monadic MLNs is polynomial in $|\mathbf{X}|$, and therefore also in the domain size $|\mathbf{D}|$.

COROLLARY 6. *Inference in monadic MLNs is domain-lifted.*

5. MARGINAL AND CONDITIONAL EXCHANGEABILITY

Many distributions are not decomposable into independent or exchangeable decompositions. Similar to conditional independence, the notion of exchangeability can be extended to conditional exchangeability. We generalize exchangeability to conditional distributions, and state the corresponding tractability guarantees.

5.1 Marginal and Conditional Decomposition

Tractability results for exchangeable decompositions on all variables under consideration also extend to subsets.

DEFINITION 7 (MARGINAL EXCHANGEABILITY). *When a subset $\mathbf{Y}$ of the variables under consideration has an exchangeable decomposition $\mathcal{Y}$, we say that $\mathbf{Y}$ is* marginally exchangeable.

This means that $\mathcal{Y}$ is still an exchangeable decomposition when considering the distribution $\Pr(\mathbf{Y}) = \sum_{\mathbf{X}\setminus\mathbf{Y}} \Pr(\mathbf{X})$.

THEOREM 7. *Suppose we are given a marginally exchangeable decomposition of $\mathbf{Y}$ with bounded width and let $\mathbf{Z} = \mathbf{X} \setminus \mathbf{Y}$. If computing $\Pr(\mathbf{y}) = \sum_{\mathbf{z}} \Pr(\mathbf{y}, \mathbf{z})$ is polynomial in $|\mathbf{X}|$ for all $\mathbf{y}$, then the complexity of MPE and marginal inference over variables $\mathbf{Y}$ is polynomial in $|\mathbf{X}|$.*

PROOF. Let T be the statistic associated with the given decomposition, and let $\mathbf{e}$ be evidence given on $\mathbf{E} \subseteq \mathbf{Y}$. Then,

$$P(\mathbf{e}) = \sum_{t \in \mathcal{T}} |S_{t,\mathbf{e}}| \cdot \sum_{\mathbf{z}} \Pr(\mathbf{y}_{t,\mathbf{e}}, \mathbf{z}), \quad \text{where} \quad \mathbf{y}_{t,\mathbf{e}} \in S_{t,\mathbf{e}}.$$

By the assumption that $\mathbf{Y}$ is marginally exchangeable and the proof of Theorem 4, we can compute $|S_{t,\mathbf{e}}|$ and an $\mathbf{y}_{t,\mathbf{e}} \in S_{t,\mathbf{e}}$ in time polynomial in $|\mathbf{X}|$. An analogous argument holds for MPE inference on $\Pr(\mathbf{Y})$. □

We now need to identify distributions $\Pr(\mathbf{Y}, \mathbf{Z})$ for which we can compute $\Pr(\mathbf{y})$ efficiently. This implies tractable probabilistic inference over $\mathbf{Y}$. Given a particular $\mathbf{y}$, we have already seen sufficient conditions: when $\mathbf{Z}$ decomposes exchangeably or independently conditioned on $\mathbf{y}$, Proposition 3 and Theorem 4 guarantee that computing $\Pr(\mathbf{y})$ is tractable. This suggests the following general notion.

Definition 8 (Conditional Decomposability). *Let $\mathbf{X}$ be a set of variables with $\mathbf{Y} \subseteq \mathbf{X}$ and $\mathbf{Z} = \mathbf{X} \setminus \mathbf{Y}$. We say that $\mathbf{Z}$ is* exchangeably (independently) decomposable given $\mathbf{Y}$ *if and only if for each assignment $\mathbf{y}$ to $\mathbf{Y}$, there exists an exchangeable (independent) decomposition $\mathcal{Z}_\mathbf{y}$ of $\mathbf{Z}$.*

Furthermore, we say that $\mathbf{Z}$ is decomposable with *bounded width* iff the width w of each $\mathcal{Z}_\mathbf{y}$ is bounded. When the decomposition can be computed in time polynomial in $|\mathbf{X}|$ for all $\mathbf{y}$, we say that $\mathbf{Z}$ is *efficiently decomposable*.

Theorem 8. *Let $\mathbf{X}$ be a set of variables with $\mathbf{Y} \subseteq \mathbf{X}$ and $\mathbf{Z} = \mathbf{X} \setminus \mathbf{Y}$. Suppose we are given a marginally exchangeable decomposition of $\mathbf{Y}$ with bounded width. Suppose further that $\mathbf{Z}$ is efficiently (exchangeably or independently) decomposable given $\mathbf{Y}$ with bounded width. If we can compute $\Pr(\mathbf{x})$ efficiently, then the complexity of MPE and marginal inference over variables $\mathbf{Y}$ is polynomial in $|\mathbf{X}|$.*

Proof. Following Theorem 7, we only need to show that we can compute $\Pr(\mathbf{y}) = \sum_\mathbf{z} \Pr(\mathbf{y}, \mathbf{z})$ in time polynomial in $|\mathbf{X}|$ for all $\mathbf{y}$. When $\mathbf{Z}$ is exchangeably decomposable given $\mathbf{Y}$, this follows from constructing $\mathcal{Z}_\mathbf{y}$ and employing the arguments made in the proof of Theorem 4. The case when $\mathbf{Z}$ is independently decomposable is analogous. □

Theorems 7 and 8 are powerful results and allow us to identify numerous probabilistic models for which inference is tractable. For instance, we will prove liftability results for Markov logic networks. However, we are only at the beginning of leveraging these tractability results to their fullest extent. Especially Theorem 7 is widely applicable because the computation of $\sum_\mathbf{z} \Pr(\mathbf{y}, \mathbf{z})$ can be tractable for many reasons. For instance, conditioned on the variables $\mathbf{Y}$, the distributions $\Pr(\mathbf{y}, \mathbf{Z})$ could be bounded treewidth graphical models, such as tree-structured Markov networks. Tractability for $\Pr(\mathbf{Y})$ follows immediately from Theorem 7.

Since Theorem 7 only speaks to the tractability of querying variables in $\mathbf{Y}$, there is the question of when we can also efficiently query the variables $\mathbf{Z}$. Results from the lifted inference literature may provide a solution by bounding or approximating queries and evidence that includes $\mathbf{Z}$ to maintain marginal exchangeability [32]. The next section shows that certain restricted situations permit tractable inference on the variables in $\mathbf{Z}$.

5.2 Markov Logic Case Study

Let us again consider the MLN

$$1.3 \quad \texttt{Smokes}(\mathsf{x}) \Rightarrow \texttt{Cancer}(\mathsf{x})$$
$$1.5 \quad \texttt{Smokes}(\mathsf{x}) \wedge \texttt{Friends}(\mathsf{x}, \mathsf{y}) \Rightarrow \texttt{Smokes}(\mathsf{y})$$

having the marginally exchangeable decomposition

$$\mathcal{Y} = \{\{\texttt{Smokes}(\mathsf{A})\}, \{\texttt{Smokes}(\mathsf{B})\}, \{\texttt{Smokes}(\mathsf{C})\}, \dots\}$$

whose width is one. To intuitively see why this decomposition is marginally exchangeable, let us consider two states $\mathbf{y}$ and $\mathbf{y}'$ of the $\texttt{Smokes}(\mathsf{x})$ atoms in which *only* the values of two atoms, for example $\texttt{Smokes}(\mathsf{A})$ and $\texttt{Smokes}(\mathsf{B})$, are swapped. There is a symmetry of the MLNs joint distribution that swaps these atoms: the renaming automorphism [3; 20] that swaps constants A and B in all atoms. For marginal exchangeability, we need that $\sum_\mathbf{z} \Pr(\mathbf{y}, \mathbf{z}) = \sum_\mathbf{z} \Pr(\mathbf{y}', \mathbf{z})$. But this holds since the renaming automorphism is an automorphism of the set of states $\{\mathbf{yz} \mid \mathbf{z} \in \mathcal{D}_\mathbf{Z}\}$ – for every $\mathbf{y}$, $\mathbf{y}'$, and $\mathbf{z}$ there exists an automorphism that maps $\mathbf{yz}$ to $\mathbf{y}'\mathbf{z}'$ with $\Pr(\mathbf{y}, \mathbf{z}) = \Pr(\mathbf{y}', \mathbf{z}')$.

The given MLN has several marginally exchangeable decompositions, with the most general one being

$$\begin{aligned}\mathcal{Y} = \{&\{\texttt{Smokes}(\mathsf{A}), \texttt{Cancer}(\mathsf{A}), \texttt{Friends}(\mathsf{A}, \mathsf{A})\},\\ &\{\texttt{Smokes}(\mathsf{B}), \texttt{Cancer}(\mathsf{B}), \texttt{Friends}(\mathsf{B}, \mathsf{B})\},\\ &\{\texttt{Smokes}(\mathsf{C}), \texttt{Cancer}(\mathsf{C}), \texttt{Friends}(\mathsf{C}, \mathsf{C})\}, \dots\}.\end{aligned}$$

For that decomposition, the remaining $\mathbf{Z}$ variables

$$\{\texttt{Friends}(\mathsf{A}, \mathsf{B}), \texttt{Friends}(\mathsf{B}, \mathsf{A}), \texttt{Friends}(\mathsf{A}, \mathsf{C}), \dots\}$$

are independently decomposable given $\mathbf{Y}$. The $\mathbf{Z}$ variables appear at most once in any formula. In a probabilistic graphical model representation, evidence on the $\mathbf{Y}$ variables would therefore decompose the graph into independent components. Thus, it follows from Theorem 8 that we can efficiently answer any query over the variables in $\mathbf{Y}$.

This insight generalizes to a large class of MLNs, called the *two-variable fragment*. It consists of all MLNs whose formulas contain at most two logical variables.

Theorem 9. *In a two-variable fragment MLN, let $\mathbf{Y}$ and $\mathbf{Z}$ be the ground atoms with one and two distinct arguments respectively. Then there exists a marginally exchangeable decomposition of $\mathbf{Y}$, and $\mathbf{Z}$ is efficiently independently decomposable given $\mathbf{Y}$. Each decomposition's width is at most twice the number of predicates.*

The proof of Theorem 9 is rather technical and we refer the reader to the appendix for a detailed proof. It now follows from Theorems 8 and 9 that the complexity of inference over the unary atoms in the two-variable fragment is polynomial in the domain size $|\mathbf{D}|$.

What happens if our query involves variables from $\mathbf{Z}$ – the binary atoms? It is known in the lifted inference literature that we cannot expect efficient inference of general queries that involve the binary atoms. Assignments to the $\mathbf{Z}$ variables break symmetries and therefore break marginal exchangeability. This causes inference to become #P-hard as a function of the query [33]. Nevertheless, if we bound the

number of binary atoms involved in the query, we can use the developed theory to show a general liftability result.

THEOREM 10. *For any MLN in the two-variable fragment, MPE and marginal inference over the unary atoms and a bounded number of binary atoms is domain-lifted.*

This corresponds to one of the strongest known theoretical results in the lifted inference literature [11]. We refer the interested reader to the appendix for the proof. A consequence of Theorem 10 is that we can efficiently compute *all single marginals* in the two-variable fragment, given arbitrary evidence on the unary atoms.

6. DISCUSSION AND CONCLUSION

We conjecture that the concept of (partial) exchangeability has potential to contribute to a deeper understanding of tractable probabilistic models. The important role conditional independence plays in the research field of graphical models is evidence for this hypothesis. Similar to conditional independence, it might be possible to develop a theory of exchangeability that mirrors that of independence. For instance, there might be a (graphical) structural representations of particular types of partial exchangeability and corresponding logical axiomatizations [24]. Moreover, it would be interesting to develop graphical models with exchangeability and independence, and notions like d-separation to detect marginal exchangeability and conditional decomposability from a structural representation. The first author has taken steps in this direction by introducing exchangeable variable models, a class of (non-relational) probabilistic models based on finite partial exchangeability [**?**].

Recently, there has been considerable interest in computing and exploiting the automorphisms of graphical models [21; 3]. There are several interesting connections between automorphisms, exchangeability, and lifted inference [20]. Moreover, there are several group theoretical algorithms that one could apply to the automorphism groups to discover the structure of exchangeable variable decompositions from the structure of the graphical models. Since we presently only exploit renaming automorphisms, there is a potential for tractable inference in MLNs that goes beyond what is known in the lifted inference literature.

Partial exchangeability is related to collective graphical models [28] (CGMs) and cardinality-based potentials [**?**] as these models also operate on sufficient statistics. However, probabilistic inference for CGMs is not tractable and there are no theoretical results that identify tractable CGMs models. The presented work may help to identify such situations. The presented theory generalizes the statistics of cardinality-based potentials.

Lifted Inference and Exchangeability

Our case studies identified a deep connection between lifted probabilistic inference and the concepts of partial, marginal and conditional exchangeability. In this new context, it appears that exact lifted inference algorithms [6; 18; 12; 34; 10; 29] can all be understood as performing essentially three steps: (i) construct a sufficient statistic $T(\mathbf{x})$, (ii) generate all possible values of the sufficient statistic, and (iii) count suborbit sizes for a given statistic. For an example of (i), we can show that a compiled first-order circuit [34] or the trace of probabilistic theorem proving [10] encode a sufficient statistic in their existential quantifier and splitting nodes. Steps (ii) and (iii) are manifested in all these algorithms through summations and binomial coefficients.

Between Corollary 6 and Theorem 10, we have re-proven almost the entire range of liftability results from the lifted inference literature [11] within the exchangeability framework, and extended these to MPE inference. There is an essential difference though: liftability results make assumptions about the syntax (e.g., MLNs), whereas our exchangeability theorems apply to all distributions. We expect Theorem 8 to be used to show liftability, and more general tractability results for many other representation languages, including but not limited to the large number of statistical relational languages that have been proposed [9; 5].

Acknowledgments

This work was partially supported by ONR grants #N00014-12-1-0423, #N00014-13-1-0720, and #N00014-12-1-0312; NSF grants #IIS-1118122 and #IIS-0916161; ARO grant #W911NF-08-1-0242; AFRL contract #FA8750-13-2-0019; the Research Foundation-Flanders (FWO-Vlaanderen); and a Google research award to MN.

APPENDIX

A. CONTINUED PROOF OF THEOREM 4

PROOF. To prove statements (b) and (c), we need to represent partial assignments $\mathbf{E} = \mathbf{e}$ with $\mathbf{E} \subseteq \mathbf{X}$. The partial assignments decompose into partial assignments $\mathbf{E}_1 = \mathbf{e}_1, \ldots, \mathbf{E}_k = \mathbf{e}_k$ in accordance with $\mathcal{X}$. Each $\mathbf{e}_\ell$ corresponds to a string $m \in \{0, 1, *\}^w$ where characters 0 and 1 encode assignments to variables in $\mathbf{E}_\ell$ and $*$ encodes an unassigned variable in $\mathbf{X}_\ell - \mathbf{E}_\ell$. In this case, we say that $\mathbf{e}_\ell$ is of type m. Please note that there are 2^w distinct b and 3^w distinct m. We say that $\mathbf{x}$ *agrees with* $\mathbf{e}$, denoted by $\mathbf{x} \sim \mathbf{e}$, if and only if their shared variables have identical assignments.

A *completion* $\mathtt{c}$ *of* $\mathbf{e}$ *to* $\mathbf{x}$ is a bijection $\mathtt{c} : \{\mathbf{e}_1, ..., \mathbf{e}_k\} \to \{\mathbf{x}_1, ..., \mathbf{x}_k\}$ such that $\mathtt{c}(\mathbf{e}_i) = \mathbf{x}_j$ implies $\mathbf{e}_i \sim \mathbf{x}_j$. Every completion corresponds to a unique way to assign elements in $\{0, 1\}$ to unassigned variables so as to turn the partial assignment $\mathbf{e}$ into the full assignment $\mathbf{x}$.

Let $t = (c_1, ..., c_{2^w}) \in \mathcal{T}$, let $\mathbf{E} \subseteq \mathbf{X}$, and let $\mathbf{e} \in \mathcal{D}_\mathbf{E}$. Moreover, let $d_j = \sum_{\ell=1}^{k} [[\mathbf{e}_\ell = m_j]]$ for each $m_j \in \{0, 1, *\}^w$. Consider the set of matrices

$$\mathcal{A}_{t,\mathbf{e}} = \Big\{ A \in \mathbb{M}(2^w, 3^w) \;\Big|\; \sum_i a_{i,j} = d_j, \sum_j a_{i,j} = c_i \text{ and } a_{i,j} = 0 \text{ if } b_i \not\sim m_j \Big\}.$$

Every $A \in \mathcal{A}_{t,\mathbf{e}}$ represents a set of completions from $\mathbf{e}$ to $\mathbf{x}$ for which $T(\mathbf{x}) = t$. The value $a_{i,j}$ indicates that each completion represented by A maps $a_{i,j}$ elements in $\{\mathbf{e}_1, ..., \mathbf{e}_k\}$ of type m_j to $a_{i,j}$ elements in $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ of type b_i. We write $\gamma(A)$ for the set of completions represented by A.

We have to prove the following statements

1. For every $A \in \mathcal{A}_{t,\mathbf{e}}$ and every $\mathsf{c} \in \gamma(A)$ there exists an $\mathbf{x} \in \mathcal{D}_{\mathbf{X}}$ with $T(\mathbf{x}) = t$ and c is a completion of $\mathbf{e}$ to $\mathbf{x}$;
2. For every $\mathbf{x} \in \mathcal{D}_{\mathbf{X}}$ with $T(\mathbf{x}) = t$ and every completion c of $\mathbf{e}$ to $\mathbf{x}$ there exists an $A \in \mathcal{A}_{t,\mathbf{e}}$ such that $\mathsf{c} \in \gamma(A)$;
3. For all $A, A' \in \mathcal{A}_{t,\mathbf{e}}$ with $A \neq A'$ we have that $\gamma(A) \cap \gamma(A') = \emptyset$;
4. For every $A \in \mathcal{A}_{t,\mathbf{e}}$, we can efficiently compute $|\gamma(A)|$, the size of the set of completions represented by A.

To prove statement (1), let $A \in \mathcal{A}_{t,\mathbf{e}}$ and $\mathsf{c} \in \gamma(A)$. By the definition of $\mathcal{A}_{t,\mathbf{e}}$, c maps $a_{i,j}$ elements in $\{\mathbf{e}_1, ..., \mathbf{e}_k\}$ of type m_j to $a_{i,j}$ elements in $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ of type b_i. By the conditions $\sum_i a_{i,j} = d_j$ and $\sum_j a_{i,j} = c_i$ of the definition of $\mathcal{A}_{t,\mathbf{e}}$ we have that c is a bijection. By the condition $a_{i,j} = 0$ if $b_i \not\sim m_j$ of the definition of $\mathcal{A}_{t,\mathbf{e}}$ we have that $\mathsf{c}(\mathbf{e}_i) = \mathbf{x}_j$ implies $\mathbf{e}_i \sim \mathbf{x}_j$ and, therefore, $\mathbf{e} \sim \mathbf{x}$. Hence, c is a completion. Moreover, c completes $\mathbf{e}$ to an $\mathbf{x}$ with $\sum_{\ell=1}^{k} [[\mathbf{x}_\ell = b_i]] = \sum_j a_{i,j} = c_i$ by the definition of $\mathcal{A}_{t,\mathbf{e}}$. Hence, $T(\mathbf{x}) = t$.

To prove statement (2), let $\mathbf{x} \in \mathcal{D}_{\mathbf{X}}$ with $T(\mathbf{x}) = t$ and let c be a completion of $\mathbf{e}$ to $\mathbf{x}$. We construct an A with $\mathsf{c} \in \gamma(A)$ as follows. Since c is a completion we have that $\mathsf{c}(\mathbf{e}_i) = \mathbf{x}_j$ implies $\mathbf{e}_i \sim \mathbf{x}_j$ and, hence, we set $a_{i,j} = 0$ if $b_i \not\sim m_j$. For all other entries in A we set $a_{i,j} = |\{\mathbf{e}_j \mid \mathsf{c}(\mathbf{e}_j) = \mathbf{x}_i\}|$. Since c is surjective, we have that $\sum_j a_{i,j} = c_i$ and since c is injective, we have that $\sum_i a_{i,j} = d_j$. Hence, $A \in \mathcal{A}_{t,\mathbf{e}}$.

To prove statement (3), let $A, A' \in \mathcal{A}_{t,\mathbf{e}}$ with $A \neq A'$. Since $A \neq A'$ we have that there exist i, j such that, without loss of generality, $a_{i,j} < a'_{i,j}$. Hence, every $\mathsf{c} \in A$ maps fewer elements of type m_j to elements of type b_i than every $\mathsf{c}' \in A'$. Hence, $\mathsf{c} \neq \mathsf{c}'$ for every $\mathsf{c} \in A$ and every $\mathsf{c}' \in A'$.

To prove statement (4), let $A \in \mathcal{A}_{t,\mathbf{e}}$. Every $\mathsf{c} \in A$ maps $a_{i,j}$ elements in $\{\mathbf{e}_1, ..., \mathbf{e}_k\}$ of type m_j to $a_{i,j}$ elements in $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ of type b_i. Hence, the size of the set of completions represented by A is, for each $1 \leq j \leq 3^w$, the number of different ways to place $a_{i,j}$ balls of color i, $1 \leq i \leq 2^w$, into d_j urns. Hence,

$$|\gamma(A)| = \prod_{j=1}^{3^w} \prod_{i=1}^{2^w} \binom{d_j - \sum_{q=1}^{i-1} a_{i,q}}{a_{i,j}}.$$

From the statements (1)-(4) we can conclude that

$$|S_{t,\mathbf{e}}| = \sum_{A \in \mathcal{A}_{t,\mathbf{e}}} |\gamma(A)|.$$

This allows us to prove (b) and (c). We can construct $\mathcal{A}_{t,\mathbf{e}}$ in time polynomial in n as follows. There are $2^w 3^w$ entries in a $2^w \times 3^w$ matrix and each entry has at most k different values. Hence, we can enumerate all $k^{6^w} \leq n^{6^w}$ possible matrices $A \in \mathbb{M}(2^w, 3^w)$. We simply select those A for which the conditions in the definition of $\mathcal{A}_{t,\mathbf{e}}$ hold. For one $A \in \mathcal{A}_{t,\mathbf{e}}$ we can efficiently construct one c and the $\mathbf{x}$ that it completes $\mathbf{e}$ to. This proves (b). Finally, we compute $|S_{t,\mathbf{e}}|$. This proves (c). □

B. PROOF OF THEOREM 5

PROOF. Let $P_1, ..., P_N$ be the N unary predicates of a given MLN and let $\mathbf{D} = \{1, ..., k\}$ be the domain. After grounding, there are k ground atoms per predicate. We write $P_j(i)$ to denote the ground atom that resulted from instantiating predicate P_j with domain element i. Let $\mathcal{X} = \{\mathbf{X}_1, ..., \mathbf{X}_k\}$ be a decomposition of the set of ground atoms with $\mathbf{X}_i = \{P_1(i), P_2(i), ..., P_N(i)\}$ for every $1 \leq i \leq k$. A renaming automorphism [3; 20] is a permutation of the ground atoms that results from a permutation of the domain elements. The joint distribution over all ground atoms remains invariant under these permutations. Consider the permutation of ground atoms that results from swapping two domain elements $i \leftrightarrow i'$. This permutation acting on the set of ground atoms permutes the components $\mathbf{X}_i$ and $\mathbf{X}_{i'}$ and leaves all other components invariant. Since this is possible for each pair $i, i' \in \{1, ..., k\}$ it follows that the decomposition $\mathcal{X}$ is exchangeable. □

C. PROOF OF THEOREM 9

PROOF. Let $P_1, ..., P_M$ be the M unary predicates and let $Q_1, ..., Q_N$ be the N binary predicates of a given MLN and let $\mathbf{D} = \{1, ..., k\}$ be the domain. After grounding, there are k ground atoms per unary and k^2 ground atoms per binary predicate. We write $P_\ell(i)$ to denote the ground atom that resulted from instantiating unary predicate P_ℓ with domain element i and $Q_\ell(i, j)$ to denote the ground atom that resulted from instantiating binary predicate Q_ℓ with domain elements i and j.

Let $\mathbf{X}$ be the set of all ground atoms, let $\mathbf{Y} = \{P_\ell(i) \mid 1 \leq i \leq k, 1 \leq \ell \leq M\} \cup \{Q_\ell(i, i) \mid 1 \leq i \leq k, 1 \leq \ell \leq N\}$ and let $\mathbf{Z} = \mathbf{X} - \mathbf{Y}$. Moreover, let $\mathcal{Y} = \{\mathbf{Y}_1, ..., \mathbf{Y}_k\}$ with $\mathbf{Y}_i = \{P_\ell(i) \mid 1 \leq \ell \leq M\} \cup \{Q_\ell(i, i) \mid 1 \leq \ell \leq N\}$. We can make the same arguments as in the proof of Theorem 5 to show that $\mathcal{Y}$ is exchangeable.

Now, we prove that the variables $\mathbf{Z}$ are independently decomposable given $\mathbf{Y}$. Let $\mathbf{Z}_{i,j} = \{Q_1(i, j), ..., Q_N(i, j), Q_1(j, i), ..., Q_N(j, i)\}$ for all $1 \leq i < j \leq k$. Now, let f be any ground formula and let G be the set of ground atoms occurring in both f and $\mathbf{Z}$. Then, either $G \subseteq \mathbf{Z}_{i,j}$ or $G \cap \mathbf{Z}_{i,j} = \emptyset$ since every formula in the MLN has at most two variables. Hence, $\{\mathbf{Z}_{i,j} \mid 1 \leq i < j \leq k\}$ is a decomposition of $\mathbf{Z}$ with $\binom{k}{2}$ components, width $2N$, and $\Pr(\mathbf{Z}, \mathbf{y})$ factorizes as

$$\Pr(\mathbf{Z}, \mathbf{y}) = \prod_{\substack{i,j \\ i<j}} Q_{i,j}(\mathbf{Z}_{i,j}, \mathbf{y}).$$

By the properties of MLNs, we have that the $Q_{i,j}(\mathbf{Z}_{i,j}, \mathbf{y})$ are computable in time exponential in the width of the decomposition but polynomial in k. □

D. PROOF OF THEOREM 10

PROOF. Suppose that the query $\mathbf{e}$ contains a bounded number of binary atoms whose arguments are constants from the set K. Consider the set of variables $\mathbf{Q}$ consisting of all unary atoms whose argument comes from K, and all binary atoms whose arguments both come from K. The unary atoms in $\mathbf{Q}$ are no long marginally exchangeable, because their arguments can appear asymmetrically in $\mathbf{e}$. We can now answer the query by simply enumerating all states $\mathbf{q}$ of $\mathbf{Q}$ and performing inference in each $\Pr(\mathbf{Y}, \mathbf{Z}, \mathbf{q})$ separately, were all variables $\mathbf{Y}$ have again become marginally exchangeable, and all variables $\mathbf{Z}$ have become independently decomposable given $\mathbf{Y}$. The construction of $\mathbf{Y}$ and $\mathbf{Z}$ is similar to the proof of Theorem 9, except that some additional binary atoms are now in $\mathbf{Y}$ instead of $\mathbf{Z}$. These atoms have one argument in K, and one not in K, and are treated as unary. When we bound the number of binary atoms in the query, the size of $\mathbf{Q}$ will not be a function of the domain size, and enumerating over all states $\mathbf{q}$ is domain-lifted. □

References

[1] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian networks. In *Proceedings of the 12th Conference on Uncertainty in artificial intelligence (UAI)*, pages 115–123, 1996.

[2] H. Bui, T. Huynh, and R. de Salvo Braz. Exact lifted inference with distinct soft evidence on every object. In *Proceedings of the 26th Conference on Artificial Intelligence (AAAI)*, 2012.

[3] H. Bui, T. Huynh, and S. Riedel. Automorphism groups of graphical models and lifted variational inference. 2012.

[4] A. Darwiche. *Modeling and reasoning with Bayesian networks*. Cambridge University Press, 2009.

[5] L. De Raedt, P. Frasconi, K. Kersting, and S. Muggleton, editors. *Probabilistic inductive logic programming: theory and applications*. Springer-Verlag, Berlin, Heidelberg, 2008.

[6] R. de Salvo Braz, E. Amir, and D. Roth. Lifted first-order probabilistic inference. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1319–1325, 2005.

[7] P. Diaconis and D. Freedman. De Finetti's generalizations of exchangeability. In *Studies in Inductive Logic and Probability*, volume II. 1980.

[8] P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8(4):745–764, 1980.

[9] L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.

[10] V. Gogate and P. Domingos. Probabilistic theorem proving. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 256–265, 2011.

[11] M. Jaeger and G. Van den Broeck. Liftability of probabilistic inference: Upper and lower bounds. In *Proceedings of the 2nd International Workshop on Statistical Relational AI*, 2012.

[12] A. Jha, V. Gogate, A. Meliou, and D. Suciu. Lifted inference seen from the other side: The tractable features. In *Proceedings of the 24th Conference on Neural Information Processing Systems (NIPS)*, 2010.

[13] K. Kersting. Lifted probabilistic inference. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, 2012.

[14] D. Koller and N. Friedman. *Probabilistic Graphical Models*. The MIT Press, 2009.

[15] S. L. Lauritzen. *Extremal families and systems of sufficient statistics*. Lecture notes in statistics. Springer-Verlag, 1988.

[16] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[17] S. L. Lauritzen, O. E. Barndorff-Nielsen, A. P. Dawid, P. Diaconis, and S. Johansen. Extreme point models in statistics. *Scandinavian Journal of Statistics*, 11(2), 1984.

[18] B. Milch, L. Zettlemoyer, K. Kersting, M. Haimes, and L. Kaelbling. Lifted probabilistic inference with counting formulas. *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pages 1062–1068, 2008.

[19] M. Mladenov and K. Kersting. Lifted inference via k-locality. In *Proceedings of the 3rd International Workshop on Statistical Relational AI*, 2013.

[20] M. Niepert. Lifted probabilistic inference: An MCMC perspective. In *Proceedings of StaRAI*, 2012.

[21] M. Niepert. Markov chains on orbits of permutation groups. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2012.

[22] M. Niepert. Symmetry-aware marginal density estimation. In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI)*, 2013.

[23] J. Noessner, M. Niepert, and H. Stuckenschmidt. RockIt: Exploiting Parallelism and Symmetry for MAP Inference in Statistical Relational Models. In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI)*, 2013.

[24] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[25] D. Poole. First-order probabilistic inference. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 985–991, 2003.

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

[27] N. Robertson and P. Seymour. Graph minors. II. Algorithmic aspects of tree-width. *Journal of algorithms*, 7(3):309–322, 1986.

[28] D. Sheldon and T. Dietterich. Collective graphical models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1161–1169. 2011.

[29] N. Taghipour, D. Fierens, J. Davis, and H. Blockeel. Lifted variable elimination with arbitrary constraints. In *Proceedings of the fifteenth international conference on Artificial Intelligence and Statistics*, volume 22, pages 1194–1202, 2012.

[30] N. Taghipour, D. Fierens, G. Van den Broeck, J. Davis, and H. Blockeel. Completeness results for lifted variable elimination. In *Proceedings of the 16th Conference on Artificial Intelligence and Statistics*, pages 572–580, 2013.

[31] G. Van den Broeck. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Advances in Neural Information Processing Systems 24 (NIPS),*, pages 1386–1394, 2011.

[32] G. Van den Broeck and A. Darwiche. On the complexity and approximation of binary evidence in lifted inference. In *Advances in Neural Information Processing Systems 26 (NIPS)*, 2013.

[33] G. Van den Broeck and J. Davis. Conditioning in first-order knowledge compilation and lifted probabilistic inference. In *Proceedings of the 26th Conference on Artificial Intelligence (AAAI)*, 2012.

[34] G. Van den Broeck, N. Taghipour, W. Meert, J. Davis, and L. De Raedt. Lifted probabilistic inference by first-order knowledge compilation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2178–2185, 2011.