10 Lifted Markov Chain Monte Carlo

Mathias Niepert and Guy Van den Broeck

Abstract.
This chapter presents an approach to utilize exact and approximate symmetries in probabilistic graphical models during Markov chain Monte Carlo (MCMC) inference. We discuss permutation groups representing the symmetries of graphical models and how to compute them. Next, we introduce orbital Markov chains, a family of lifted Markov chains leveraging model symmetries to reduce mixing times. Unfortunately, the majority of real-world graphical models is asymmetric. This is even the case for relational representations when evidence is given. Therefore, we extend lifted MCMC to instead utilize approximate symmetries. Lifted MCMC leads to improved probability estimates while remaining unbiased. Experiments demonstrate that the approach outperforms existing MCMC algorithms.

10.1 Introduction

This chapter describes the use of group theoretical concepts and algorithms to perform Markov chain Monte Carlo sampling in probabilistic models. Since relational models often exhibit strong topological symmetries, permutation groups offer a compact and well-understood representation. Moreover, numerous efficient group theoretical algorithms are implemented in comprehensive open-source group algebra frameworks such as GAP (GAP).

Symmetries on different syntactical levels of statistical relational formalism ultimately lead to symmetries in the space of joint variable assignments. This space of possible assignments corresponds to the state space of Monte Carlo Markov chains such as the Gibbs sampler that are often used for approximate probabilistic inference. Since the permutation group modeling the symmetries induces a partition (the so-called orbit partition) on the state space of these Markov chains, we investigate whether this can be exploited for more efficient MCMC approaches to probabilistic inference. The basic idea is that *lifted* Markov chains implicitly or explicitly operate on the partition of the state space instead of the space of individual assignments. We describe orbital Markov chains, which are derived from an existing Markov chain so as to leverage the symmetries in the underlying model. Under mild conditions, orbital Markov chains have the same convergence properties as chains operating on the state space partition without the need to explicitly compute this partition.

While lifted inference algorithms perform well for highly symmetric graphical models, they depend heavily on the presence of symmetries and perform worse for asymmetric models due to their computational overhead. This is especially unfortunate as numerous real-world graphical models are not symmetric. To bring the achievements of the lifted inference community to the mainstream of machine learning and uncertain reasoning it is crucial to explore ways to apply ideas from the lifted inference literature to inference problems in asymmetric graphical models. This chapter further describes a lifted inference algorithm for asymmetric graphical models. It uses a symmetric approximation of the original model to compute a proposal distribution for a Metropolis-Hastings chain. The approach combines a base MCMC algorithm such as the Gibbs sampler with the Metropolis chain that performs jumps in the approximate symmetric model, while producing unbiased probability estimates.

We conducted several experiments verifying that orbital Markov chains converge faster to the true distribution than state of the art Markov chains. We also conduct experiments where lifted inference is applied to graphical models with no exact symmetries and no color-passing symmetries, and where every random variable has distinct soft evidence. Yet, we are able to show improved probability estimates while remaining unbiased.

10.2 Background

We first recall basic concepts of group theory and finite Markov chains both of which are crucial for understanding this chapter.

10.2.1 Group Theory

A **symmetry** of a discrete object is a structure-preserving bijection on its components. A natural way to represent symmetries are **permutation groups**. A group is an algebraic structure $(\mathfrak{G}, \circ)$, where $\mathfrak{G}$ is a set closed under a binary associative operation $\circ$ with an identity element and a unique inverse for each element. We often write $\mathfrak{G}$ rather than $(\mathfrak{G}, \circ)$.

A permutation group $\mathfrak{G}$ *acting on* a finite set Ω is a finite set of bijections $\mathfrak{g} : \Omega \to \Omega$ that form a group. Let Ω be a finite set and let $\mathfrak{G}$ be a permutation group acting on Ω. If $\alpha \in \Omega$ and $\mathfrak{g} \in \mathfrak{G}$ we write $\alpha^{\mathfrak{g}}$ to denote the image of α under $\mathfrak{g}$. A cycle $(\alpha_1\ \alpha_2\ ...\ \alpha_n)$ represents the permutation that maps α_1 to α_2, α_2 to α_3,..., and α_n to α_1. Every permutation can be written as a product of disjoint cycles where each element that does not occur in a cycle is understood as being mapped to itself. A generating set R of a group is a subset of the group's elements such that every element of the group can be written as a product of finitely many elements of R and their inverses.

We define a relation $\sim$ on Ω with $\alpha \sim \beta$ if and only if there is a permutation $\mathfrak{g} \in \mathfrak{G}$ such that $\alpha^{\mathfrak{g}} = \beta$. The relation partitions Ω into equivalence classes which we call **orbits**. We call this partition of Ω the orbit partition induced by $\mathfrak{G}$. We use the notation $\alpha^{\mathfrak{G}}$ to denote the orbit $\{\alpha^{\mathfrak{g}} \mid \mathfrak{g} \in \mathfrak{G}\}$ containing α. Let $f : \Omega \to \mathbb{R}$ be a function from Ω into the real numbers and let $\mathfrak{G}$ be a permutation group acting on Ω. We say that $\mathfrak{G}$ is an **automorphism group** for (Ω, f) if and only if for all $\omega \in \Omega$ and all $\mathfrak{g} \in \mathfrak{G}$, $f(\omega) = f(\omega^{\mathfrak{g}})$.

10.2.2 Finite Markov Chains

Given a finite set Ω a **Markov chain** defines a random walk $(\mathbf{x}_0, \mathbf{x}_1, ...)$ on elements of Ω with the property that the conditional distribution of $\mathbf{x}_{n+1}$ given $(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n)$ depends only on $\mathbf{x}_n$. For all $\mathbf{x}, \mathbf{y} \in \Omega$, $P(\mathbf{x} \to \mathbf{y})$ is the chain's probability to transition from $\mathbf{x}$ to $\mathbf{y}$, and $P^t(\mathbf{x} \to \mathbf{y}) = P^t_{\mathbf{x}}(\mathbf{y})$ the probability of being in state $\mathbf{y}$ after t steps if the chain starts at state $\mathbf{x}$. We often refer to the conditional probability matrix P as the *kernel* of the Markov chain. A Markov chain is *irreducible* if for all $\mathbf{x}, \mathbf{y} \in \Omega$ there exists a t such that $P^t(\mathbf{x} \to \mathbf{y}) > 0$ and *aperiodic* if for all $\mathbf{x} \in \Omega$, $\mathsf{gcd}\{t \geq 1 \mid P^t(\mathbf{x} \to \mathbf{x}) > 0\} = 1$.

Theorem 10.1 *Any irreducible and aperiodic Markov chain has exactly one* ***stationary distribution****.*

A distribution π on Ω is reversible for a Markov chain with state space Ω and transition probabilities P, if for every $\mathbf{x}, \mathbf{y} \in \Omega$

$$\pi(\mathbf{x})P(\mathbf{x} \to \mathbf{y}) = \pi(\mathbf{y})P(\mathbf{y} \to \mathbf{x}).$$

We say that a Markov chain is reversible if there exists a reversible distribution for it. The AI literature often refers to reversible Markov chains as Markov chains satisfying the detailed balance property.

Theorem 10.2 *Every reversible distribution for a Markov chain is also a stationary distribution for the chain.*

10.2.3 Markov Chain Monte Carlo

Numerous approximate inference algorithms for probabilistic graphical models draw sample points from a Markov chain whose stationary distribution is that of the probabilistic model, and use the sample points to estimate marginal probabilities. Sampling approaches of this kind are referred to as **Markov chain Monte Carlo** methods. We discuss the **Gibbs sampler**, a sampling algorithm often used in practice.

Let $\mathbf{X}$ be a finite set of random variables with probability distribution π. The Markov chain for the *Gibbs sampler* is a Markov chain $\mathcal{M} = (\mathbf{x}_0, \mathbf{x}_1, ...)$ which, being in state $\mathbf{x}_t$, performs the following steps at time $t + 1$:

1. Select a variable $X \in \mathbf{X}$ uniformly at random;
2. Sample $\mathbf{x}'_{t+1}(X)$, the value of X in the state $\mathbf{x}'_{t+1}$, according to the conditional π-distribution of X given that all other variables take their values according to $\mathbf{x}_t$; and
3. Let $\mathbf{x}'_{t+1}(Y) = \mathbf{x}_t(Y)$ for all variables $Y \in \mathbf{X} \setminus \{X\}$.

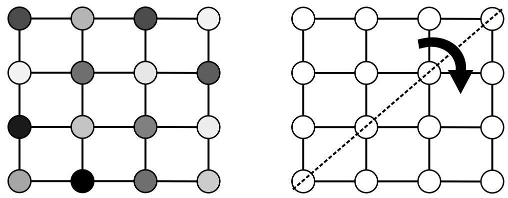


Figure 10.1: A ferromagnetic Ising model with constant interaction strength. In the presence of an external field, that is, when the variables have different unary potentials, the probabilistic model is asymmetric (a). However, the model is rendered symmetric by assuming a constant external field (b). In this case, the symmetries of the model are generated by the reflection and rotation automorphisms.

The Gibbs chain is aperiodic and has π as a stationary distribution. If the chain is irreducible, then the marginal estimates based on sample points drawn from the chain are unbiased once the chain reaches the stationary distribution.

10.3 Symmetries of Probabilistic Models

This section describes notions of symmetry in the context of probabilistic graphical models, as well as their relational extensions.

10.3.1 Graphical Model Symmetries

Symmetries of a set of random variables and graphical models have been formally defined in the lifted and symmetry-aware probabilistic inference literature with concepts from group theory (Niepert, 2013; Bui et al., 2012).

Definition 10.1 *Let* $\mathbf{X}$ *be a set of discrete random variables with distribution* π *and let* Ω *be the set of states (configurations) of* $\mathbf{X}$. *We say that a permutation group* $\mathfrak{G}$ *acting on* Ω *is an automorphism group for* $\mathbf{X}$ *if and only if for all* $\mathbf{x} \in \Omega$ *and all* $g \in \mathfrak{G}$ *we have that* $\pi(\mathbf{x}) = \pi(\mathbf{x}^g)$.

Note that the definition of an automorphism group is independent of the particular representation of the probabilistic model. For particular representations, there are efficient algorithms for computing the automorphism group. Typically, one computes the generators of the automorphism group with algorithms that derive permutation groups for colored undirected graphs such as SAUCY and NAUTY (Niepert, 2012b). Note that we do not require the automorphism group to be maximal, that is, it can be a subgroup of a different automorphism group for the same set of random variables.

Most probabilistic models are asymmetric. For instance, the Ising model which is used in numerous applications, is asymmetric if we assume an external field as it leads to different unary potentials. However, we can make the model symmetric simply by assuming a constant external field. Figure 10.1 depicts this situation. This is an example of an over-symmetric approximation of the model, which we will use later in this chapter to do lifted MCMC without biasing the probability estimates.

10.3.2 Relational Model Symmetries

Naturally, there is a close connection between the concept of symmetry and lifted inference. There are deep connections between automorphisms and the statistical notion of exchangeability (Niepert, 2012b, 2013; Bui et al., 2012), which has been used to explain the tractability of exact lifted inference algorithms (Niepert and Van den Broeck, 2014). Moreover, the (fractional) automorphisms of the graphical model representation have been related to lifted inference and exploited for more efficient inference (Niepert, 2012b; Bui et al., 2012; Noessner et al., 2013; Mladenov and Kersting, 2013). For instance, lifted belief propagation identifies and clusters indistinguishable ground atoms and features by keeping track of the messages send and received by each of the corresponding nodes in a factor graph (Singla and Domingos, 2008; Kersting et al., 2009). Bi-simulation type procedures group indistinguishable elements and, therefore, exploit symmetry in the model as well (Sen et al., 2009b). There are a number of sampling algorithms that take advantage of symmetries (Venugopal and Gogate, 2012; Gogate et al., 2012).

The algorithms in this chapter use group theory and, in particular, permutation groups to compactly represent (exact and approximate) symmetries in graphical models (Niepert, 2012b). There are several reasons to consider group theory and permutation groups a natural representation of symmetries in graphical models. First, an irredundant set of generators of a permutation group ensures exponential compression. For instance, for a set of n exchangeable binary random variables, the permutation group acting on the variables is the symmetric group on n which has $n!$ permutations. However, we only need at most $n - 1$ irredundant generators to represent this permutation group. In addition to the compact representation, group theory also provides numerous remarkably efficient algorithms for manipulating and sampling from groups. The product replacement algorithm (Celler et al., 1995), for instance, samples group elements uniformly at random with impressive performance.

Symmetry in statistical relational languages manifests itself at various syntactic levels ranging from the set of constants to the assignment space. There is often symmetry at the level of constants. In the well-known social network model (Singla and Domingos, 2008) without evidence, for example, we have that the constants are indistinguishable meaning that swapping two constants leads to an isomorphic statistical relational model. Now, the permutations on the constant level induce permutations on the level of ground atoms and formulas. From the irredundant generators of the permutation group modeling the symmetries on the constant level we can directly compute the irredundant generators of the permutation group modeling the corresponding symmetries on the ground level. Indeed, Bob $\longleftrightarrow$ Anna

smokes(Anna) $\longleftrightarrow$ smokes(Bob)

¬smokes(Anna) ∨ cancer(Peter) $\longleftrightarrow$ ¬smokes(Bob) ∨ cancer(Peter)

{..., smokes(Anna)=**1**, smokes(Bob)=**0**, ...} $\longleftrightarrow$ {..., smokes(Anna)=**0**, smokes(Bob)=**1**, ...}

Figure 10.2: Symmetry in the model is observable on different syntactical levels of the relational model. The level of constants, the level of ground atoms (variables), the level of clauses (features) and the level of possible worlds (assignments). Each permutation group acting on the set of constants induces a permutation group acting on the set of ground atoms. The latter induces a permutation group acting on the set of features. This permutation partitions (a) the variables and feature and (b) the assignment space.

it is well-known that isomorphisms between permutation groups always map irredundant generators in one group to irredundant generators in the other. However, symmetry on the ground level does not necessarily lead to symmetry on the constant level. Similarily, while symmetry on the ground level induces symmetry on the space of assignments to the random variables this is not true for the other direction. Figure 10.2 depicts the different syntactical levels on which symmetries can arise.

Niepert (2012b) describes an approach that maps weighted formulas to colored undirected graphs and applies graph automorphism algorithms to compute the symmetries of the log-linear models defined over the weighted formulas (Niepert, 2012b). The resulting permutation groups partition the (exponential) space of variable assignments when acting on it. Since the state space of MCMC approaches is identical to the assignment space of the probabilistic graphical models, we will investigate whether and to what extend the partition induced by the models' symmetries can be leveraged for more efficient MCMC algorithms.

10.4 Lifted MCMC for Symmetric Models

We have seen that symmetries on different syntactical levels of statistical relational formalism ultimately lead to symmetries in the space of joint variable assignments. Now, the space of possible variable assignments is the state space of Monte Carlo Markov chains such as the Gibbs sampler that are often used for approximate probabilistic inference. Since the permutation group modeling the symmetries induces a partition (the so-called orbit partition) on the state space of these Markov chains, we will investigate whether this can be exploited for more efficient MCMC approaches to probabilistic inference. The ba-

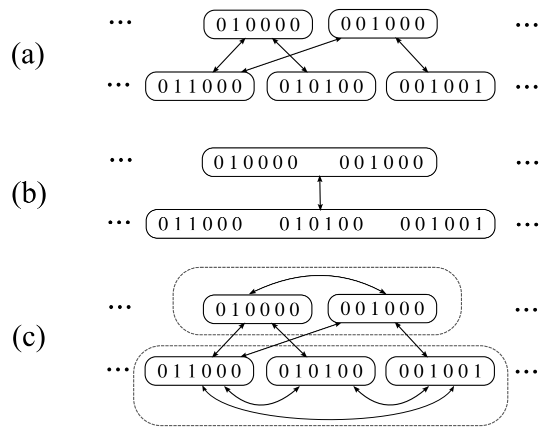


Figure 10.3: (a) A fragment of a finite state space of a Markov chain with non-zero transition probabilities indicated by directed arcs. (b) A lumping of the state space. Instead of moving between individual states, the lumped chain moves between classes of states of the original chain. (c) The benefits of lumping are also achievable by sampling uniformly at random from the implicit equivalence classes (orbits) in each step.

sic idea is that *lifted* MCMC algorithms implicitly or explicitly operate on the partition of the state space instead of the original state space.

10.4.1 Lumping

A *lumping* (also: collapsing, projection) of a Markov chains is a compression of its state space which is possible under certain conditions on the transition probabilities of the original Markov chain (Buchholz, 1994; Derisavi et al., 2003). The following definition formalizes the notion.

Definition 10.2 *Let $\mathcal{M}$ be a Markov chain with transition matrix P and state space Ω, and let $C = \{C_1, ..., C_n\}$ be a partition of the state space. If for all $C_i, C_j \in C$ and all $j', j'' \in C_j$*

$$\sum_{i' \in C_i} P(i', j') = \sum_{i' \in C_i} P(i', j'')$$

then $\mathcal{M}$ is ordinary lumpable. If, in addition, the stationary distribution has $\pi(j') = \pi(j'')$ for all $j', j'' \in C_j$ and all $C_j \in C$ then $\mathcal{M}$ is exactly lumpable.

Let $\hat{\pi}$ be the stationary distribution of the *quotient Markov chain*, that is, the exactly lumped Markov chain whose state space consists of partitions C. Then, the probability $\pi(i)$ of a state $i \in C_i \subseteq \Omega$ of the original chain can be computed as $\pi(i) = \hat{\pi}(i)/|C_i|$.

The benefit of lumping a Markov chain is the potentially much smaller state space and ultimately more rapid mixing. For instance, consider the case of n binary random variables that are exchangeable. Here, the natural choice of a partition of the state space is $\{C_0, C_1, ..., C_n\}$ where each C_i contains the states with Hamming weight i, that is, the states with i non-zeros. Please note that the C_i's are the orbits (equivalence classes) of the orbit partition of the permutation group acting on the set of states (variable assignments). Instead of 2^n states the resulting lumped Markov chain has only $n + 1$ states and mixes more rapidly than the original one. Figure 10.3 depicts (a) a fragment of a finite Markov chain with non-zero probability transitions indicated by arrows and (b) a lumped Markov chain that bundles several states of the original chain into a single one of the lumped chain.

The crucial question is whether the explicit construction of the lumped chain is computationally feasible. After all, if the computation of lumped Markov chains was intractable we would not have gained much. Unfortunately, it turns out that the explicit construction of the lumped state space is indeed intractable. Computing the coarsest lumping quotient of a Markov chain with a bi-simulation procedure is linear in the *number of non-zero probability transitions* of the chain (Derisavi et al., 2003) and, hence, in most cases exponential in the number of random variables. Moreover, other theoretical results show that special cases of the lumping problem are also intractable. The results are negative even for the important special case of partitions resulting from permutation groups acting on the state space of the Markov chains. It is known that, given a permutation group acting on the state space, merely computing the number of equivalence classes of the resulting orbit partition of the state space is a #P-complete problem (Goldberg, 2001).

We hypothesize that the intractability of the explicit construction of the lumped chain's state space is the main reason that the technique of lumping, while well-understood on a theoretical level, has not been seriously considered by communities that apply Markov chain Monte Carlo methods to large-scale applications requiring probabilistic inference. We are not aware of MCMC approaches to probabilistic reasoning that leverage the theory of lumping.

10.4.2 Orbital Markov Chains

Under certain circumstances, the *explicit* computation of the partition of the state space is not necessary to achieve the same computational gains as the lumped chain (Niepert, 2012b). The basic idea is that we only need, for each $\omega \in \Omega$, an efficient way to sample uniformly at random from $[\omega]$ the equivalence class containing ω. The product replacement algorithm (Celler et al., 1995) provides such an efficient method of sampling uniformly from the equivalence classes induced by a permutation group. This novel family of Markov chains is referred to as **orbital Markov chains** (Niepert, 2012b). An orbital Markov chain is always derived from an existing Markov chain so as to leverage the symmetries in the underlying model. In the presence of symmetries orbital Markov chains are able to perform wide-ranging transitions reducing the time until convergence. In the absence of symmetries they are equivalent to the original Markov chains. Orbital Markov chains only require a generating set of a permutation group $\mathfrak{G}$ acting on the chain's state space as additional input. These generators can be computed on a colored graph representation of the distribution at hand, or directly on the relational representation, as described in the previous section.

Let Ω be a finite set, let $\mathcal{M}' = (X'_0, X'_1, ...)$ be a Markov chain with state space Ω, let π be a stationary distribution of $\mathcal{M}'$, and let $\mathfrak{G}$ be an automorphism group for (Ω, π). The *orbital Markov chain* $\mathcal{M} = (X_0, X_1, ...)$ *for* $\mathcal{M}'$ is a Markov chain which at each integer time $t + 1$ performs the following steps:

1. Let X'_{t+1} be the state of the *original* Markov chain $\mathcal{M}'$ at time $t + 1$;
2. Sample X_{t+1}, the state of the orbital Markov chain $\mathcal{M}$ at time $t+1$, uniformly at random from ${X'}^{\mathfrak{G}}_{t+1}$, the orbit of X'_{t+1}.

The orbital Markov chain $\mathcal{M}$, therefore, runs at every time step $t \geq 1$ the original chain $\mathcal{M}'$ first and samples the state of $\mathcal{M}$ at time t uniformly at random from the orbit of the state of the original chain $\mathcal{M}'$ at time t. Figure 10.3 (c) depicts a fragment of the orbital Markov chain for the original Markov chain (a). Instead of computing the equivalence of the state space explicitely (b) novel transitions are introduced that make the chain behave *as if it was lumped*.

Given a state X_t and a permutation group $\mathfrak{G}$ orbital Markov chains sample an element from ${X_t}^{\mathfrak{G}}$, the orbit of X_t, uniformly at random. By the orbit-stabilizer theorem this is equivalent to sampling an element $\mathfrak{g} \in \mathfrak{G}$ uniformly at random and computing ${X_t}^{\mathfrak{g}}$. Sampling group elements uniformly at random is a well-researched problem (Celler et al., 1995) and computable in polynomial time in the size of the generating sets with product replacement algorithms (Pak, 2000). These algorithms are implemented in several group algebra systems such as GAP (GAP) and exhibit remarkable performance. Once initialized, product replacement algorithms can generate pseudo-random elements by performing, depending on the variant, 1 to 3 group multiplications. We could verify that the overhead of step 2 during the sampling process is indeed negligible.

The following theorem relates properties of the orbital Markov chain to those of the Markov chain it is derived from. A detailed proof can be found in the appendix.

Theorem 10.3 (Niepert (2012b)) *Let Ω be a finite set and let $\mathcal{M}'$ be a Markov chain with state space Ω and transition matrix P'. Moreover, let π be a probability distribution on Ω, let $\mathfrak{G}$ be an automorphism group for (Ω, π), and let $\mathcal{M}$ be the orbital Markov chain for $\mathcal{M}'$. Then,*

(a) if $\mathcal{M}'$ is aperiodic then $\mathcal{M}$ is also aperiodic;
(b) if $\mathcal{M}'$ is irreducible then $\mathcal{M}$ is also irreducible;
(c) if π is a reversible distribution for $\mathcal{M}'$ and, for all $\mathfrak{g} \in \mathfrak{G}$ and all $x, y \in \Omega$ we have that $P'(x, y) = P'(x^{\mathfrak{g}}, y^{\mathfrak{g}})$, then π is also a reversible and, hence, a stationary distribution for $\mathcal{M}$.

The condition in statement (c) requiring for all $\mathfrak{g} \in \mathfrak{G}$ and all $x, y \in \Omega$ that $P'(x, y) = P'(x^{\mathfrak{g}}, y^{\mathfrak{g}})$ expresses that the original Markov chain is compatible with the symmetries captured by the permutation group $\mathfrak{G}$. This weak assumption is met by all of the practical Markov chains we are aware of and, in particular, Metropolis chains and Gibbs sampler.

10.5 Lifted MCMC for Asymmetric Models

This section extends the lifted MCMC framework to construct mixtures of Markov chains where one of the chains operates on the **approximate symmetries** of the probabilistic model. The framework assumes a base Markov chain $\mathcal{M}_{\mathrm{B}}$ such as the Gibbs chain, the MC-SAT chain (Poon and Domingos, 2006), or any other MCMC algorithm. We then construct a mixture of the base chain and an Orbital Metropolis chain which exploits approximate symmetries for its proposal distribution.

10.5.1 Mixing

Two or more Markov chains can be combined by constructing mixtures and compositions of the kernels (Tierney, 1994). Let P_1 and P_2 be the kernels for two Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$ both with stationary distribution π. Given a positive probability $0 < \alpha < 1$, a *mixture* of the Markov chains is a Markov chain where, in each iteration, kernel P_1 is applied with probability α and kernel P_2 with probability $1 - \alpha$. The resulting Markov chain has π as a stationary distribution. The following result relates properties of the individual chains to properties of their mixture.

Theorem 10.4 (Tierney (1994)) *A mixture of two Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$ is irreducible and aperiodic if at least one of the chains is irreducible and aperiodic.*

For a more in-depth discussion of combining Markov chains and the application to machine learning, we refer the interested reader to an overview paper (Andrieu et al., 2003).

10.5.2 Metropolis-Hastings Chains

Before we describe the approach in more detail, let us first review Metropolis samplers. The construction of a **Metropolis-Hastings Markov chain** is a popular general procedure for designing reversible Markov chains for MCMC-based estimation of marginal probabilities. Metropolis-Hastings chains are associated with a **proposal distribution** $Q(\cdot|\mathbf{x})$ that is utilized to *propose* a move to the next state given the current state $\mathbf{x}$. The closer the proposal distribution to the distribution π to be estimated, that is, the closer $Q(\mathbf{x} \mid \mathbf{x}_t)$ to $\pi(\mathbf{x})$ for large t, the better the convergence properties of the Metropolis-Hastings chain.

We first describe the Metropolis algorithm, a special case of the Metropolis-Hastings algorithm (Häggström, 2002). Let $\mathbf{X}$ be a finite set of random variables with probability distribution π and let Ω be the set of states of the random variables. The Metropolis chain is governed by a transition graph $G = (\Omega, \mathbf{E})$ whose nodes correspond to states of the random variables. Let $\mathfrak{n}(\mathbf{x})$ be the set of neighbors of state $\mathbf{x}$ in G, that is, all states reachable from $\mathbf{x}$ with a single transition. The Metropolis chain with graph G and distribution π has transition probabilities

$$P(\mathbf{x} \to \mathbf{y}) = \begin{cases} \frac{1}{|\mathfrak{n}(\mathbf{x})|} \min\left\{\frac{\pi(\mathbf{y})|\mathfrak{n}(\mathbf{x})|}{\pi(\mathbf{x})|\mathfrak{n}(\mathbf{y})|}, 1\right\} & \text{if } x \text{ and } y \text{ are neighbors,} \\ 1 - \sum_{y' \in \mathfrak{n}(\mathbf{x})} \frac{1}{|\mathfrak{n}(\mathbf{x})|} \min\left\{\frac{\pi(\mathbf{y}')|\mathfrak{n}(\mathbf{x})|}{\pi(\mathbf{x})|\mathfrak{n}(\mathbf{y}')|}, 1\right\} & \text{if } x = y, \\ 0, & \text{otherwise.} \end{cases}$$

Being in state $\mathbf{x}_t$ of the Markov chain $\mathcal{M} = (\mathbf{x}_0, \mathbf{x}_1, ...)$, the Metropolis sampler therefore performs the following steps at time $t+1$:

1. Select a state $\mathbf{y}$ from $\mathfrak{n}(\mathbf{x}_t)$, the neighbors of $\mathbf{x}_t$, uniformly at random;
2. Let $\mathbf{x}_{t+1} = \mathbf{y}$ with probability $\min\left\{\frac{\pi(\mathbf{y})|\mathfrak{n}(\mathbf{x})|}{\pi(\mathbf{x})|\mathfrak{n}(\mathbf{y})|}, 1\right\}$;
3. Otherwise, let $\mathbf{x}_{t+1} = \mathbf{x}_t$.

Note that the proposal distribution $Q(\cdot|\mathbf{x})$ is simply the uniform distribution on the set of $\mathbf{x}$'s neighbors. It is straight-forward to show that π is a stationary distribution for the Metropolis chain by showing that π is a reversible distribution for it (Häggström, 2002).

Now, the performance of the Metropolis chain hinges on the structure of the graph G. We would like the graph structure to facilitate global moves between high probability modes, as opposed to the local moves typically performed by MCMC chains. To design such a graph structure, we take advantage of approximate symmetries in the model.

10.5.3 Orbital Metropolis Chains

This section describes a class of **orbital Metropolis chains** that move between approximate symmetries of a distribution. The approximate symmetries form an automorphism group $\mathfrak{G}$. We will discuss approaches to obtain such an automorphism group in Section 10.6. Here, we introduce a Markov chain that takes advantage of the approximate symmetries.

Given a distribution π over random variables $\mathbf{X}$ with state space Ω, and a permutation group $\mathfrak{G}$ acting on Ω, the orbital Metropolis chain $\mathcal{M}_S$ for $\mathfrak{G}$ performs the following steps:

1. Select a state $\mathbf{y}$ from $\mathbf{x}_t^{\mathfrak{G}}$, the orbit of $\mathbf{x}_t$, uniformly at random;
2. Let $\mathbf{x}_{t+1} = \mathbf{y}$ with probability $\min\left\{\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}, 1\right\}$;
3. Otherwise, let $\mathbf{x}_{t+1} = \mathbf{x}_t$.

Note that a permutation group acting on Ω partitions the states into disjoint orbits. The orbital Metropolis chain simply moves between states in the same orbit. Hence, two states in the same orbit have the same number of neighbors and, thus, the expressions cancel out in line 2 above. It is straight-forward to show that the chain $\mathcal{M}_S$ is reversible and, hence, that it has π as a stationary distribution. However, the chain is *not* irreducible as it never moves between states that are not symmetric with respect to the permutation group $\mathfrak{G}$. In the binary case, for example, it cannot reach states with a different Hamming weight from the initial state.

10.5.4 Lifted Metropolis-Hastings

To obtain an irreducible Markov chain that exploits approximate symmetries, we construct a mixture of (a) some base chain $\mathcal{M}_{\mathrm{B}}$ with stationary distribution π for which we know that it is irreducible and aperiodic; and (b) an orbital Metropolis chain $\mathcal{M}_{\mathrm{S}}$. We can prove the following theorem.

Theorem 10.5 *Let* $\mathbf{X}$ *be a set of random variables with distribution* π *and approximate automorphisms* $\mathfrak{G}$*. Moreover, let* $\mathcal{M}_{\mathrm{B}}$ *be an aperiodic and irreducible Markov chain with stationary distribution* π*, and let* $\mathcal{M}_{\mathrm{S}}$ *be the orbital Metropolis chain for* $\mathbf{X}$ *and* $\mathfrak{G}$*. The mixture of* $\mathcal{M}_{\mathrm{B}}$ *and* $\mathcal{M}_{\mathrm{S}}$ *is aperiodic, irreducible, and has* π *as its unique stationary distribution.*

The mixture of the base chain and the orbital Metropolis chain has several advantages. First, it exploits the approximate symmetries of the model which was shown to be advantageous for marginal probability estimation (Van den Broeck and Darwiche, 2013). Second, the mixture of Markov chains performs wide ranging moves via the orbital Metropolis chain, exploring the state space more efficiently and, therefore, improving the quality of the probability estimates. Figure 10.4 depicts the state space and the transition graph of (a) the Gibbs chain and (b) the mixture of the Gibbs chain and an orbital Metropolis chain. It illustrates that the mixture is able to more freely move about the state space by jumping between orbit states. For instance, moving from state 0110 to 1001 would require 4 steps of the Gibbs chain but is possible in one step with the mixture of chains. The larger the size of the automorphism groups, the more densely connected is the transition graph. Since the moves of the orbital Metropolis chain are between approximately symmetric states of the random variables, it does not suffer from the problem of most proposals being rejected. We will be able to verify this hypothesis empirically.

The general Lifted Metropolis-Hastings framework can be summarized as follows.

1. Obtain an approximate automorphism group $\mathfrak{G}$;
2. Run the following mixture of Markov chains:
   (a) With probability $0 < \alpha < 1$, apply the kernel of the base chain $\mathcal{M}_{\mathrm{B}}$;
   (b) Otherwise, apply the kernel of the orbital Metropolis chain $\mathcal{M}_{\mathrm{S}}$ for $\mathfrak{G}$.

Note that the proposed approach is a generalization of lifted MCMC for symmetric models, as described in the previous section, essentially using it as a subroutine, and that all MH proposals are accepted if $\mathfrak{G}$ is an exact automorphism group of the original model. Moreover, note that the framework allows one to combine multiple orbital Metropolis chains with a base chain.

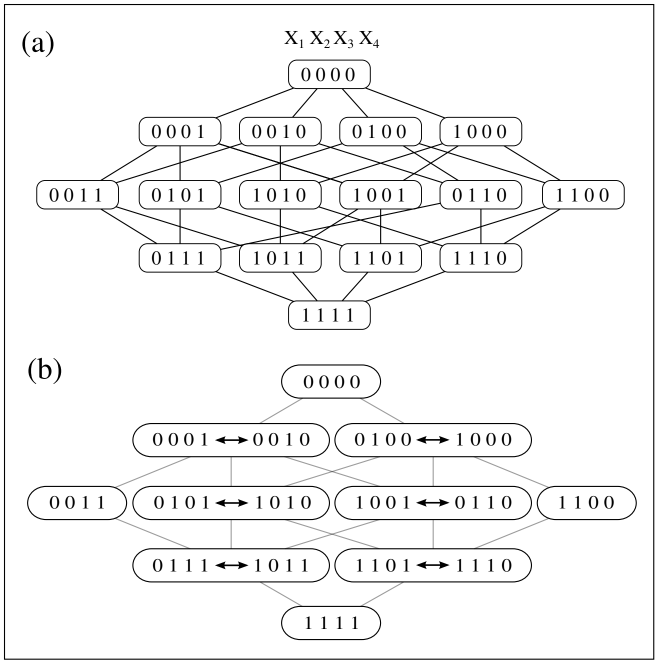


Figure 10.4: The state space (self-arcs are omitted) of (a) the Gibbs chain for four binary random variables and (b) the orbit partition of its state space induced by the permutation group generated by the permutation $(X_1\ X_2)(X_3\ X_4)$. The permutations are approximate symmetries, derived from an over-symmetric approximation of the original model. The Gibbs chain proposes moves to states whose Hamming distance to the current state is at most 1. The orbital Metropolis chain, on the other hand, proposes moves between orbit elements which have a Hamming distance of up to 4. The mixture of the two chains leads to faster convergence while maintaining an unbiased stationary distribution.

10.6 Approximate Symmetries

The Lifted Metropolis-Hastings algorithm assumes that a permutation group $\mathfrak{G}$ is given, representing the approximate symmetries. We now discuss several approaches to the computation of such an automorphism group. While it is not possible to go into technical detail here, we will provide pointers to the relevant literature.

There exist several techniques to compute the *exact symmetries* of a graphical model and construct $\mathfrak{G}$; see (Niepert, 2012b; Bui et al., 2012). The color refinement algorithm is also well-studied in lifted inference (Kersting et al., 2014). It can find (exact) orbits of random variables for a slightly weaker notion of symmetry, called fractional automorphism. These techniques all require some form of exact symmetry to be present in the model.

Detecting *approximate symmetries* is a problem that is largely open. One key idea is that of an *over-symmetric approximations* (OSAs) (Van den Broeck and Darwiche, 2013). Such approximations are derived from the original model by rendering the model more symmetric. After the computation of an over-symmetric model, we can apply existing tools for exact symmetry detection. Indeed, the exact symmetries of an approximate model are approximate symmetries of the exact model. These symmetrization techniques are indispensable to our algorithm.

10.6.1 Relational Symmetrization

Existing symmetrization techniques operate on relational representations, such as Markov logic networks (MLNs). Relational models have numerous symmetries. For example, swapping the web pages A and B in a web page classification model does not change the MLN. This permutation of constants induces a permutations of random variables (e.g., between $\mathtt{Page}(\mathsf{A}, \mathsf{Faculty})$ and $\mathtt{Page}(\mathsf{B}, \mathsf{Faculty})$). Unfortunately, hard and soft evidence breaks symmetries, even in highly symmetric relational models (Van den Broeck and Darwiche, 2013). When the variables $\mathtt{Page}(\mathsf{A}, \mathsf{Faculty})$ and $\mathtt{Page}(\mathsf{B}, \mathsf{Faculty})$ get assigned distinct soft evidence, the symmetry between A and B is removed, and lifted inference breaks down.[20] Similarly, when the `Link` relation is given, its graph is unlikely to be symmetric (Erdős and Rényi, 1963), which in turn breaks the symmetries in the MLN. These observations motivated research on OSAs. Van den Broeck and Darwiche (2013) propose to approximate binary relations, such as `Link`, by a low-rank Boolean matrix factorization. Venugopal and Gogate (2014a) cluster the constants in the domain of the MLN. Singla et al. (2014) present a message-passing approach to clustering similar constants.

10.6.2 Propositional Symmetrization

A key property of our LMH algorithm is that it operates at the propositional level, regardless of how the graphical model was generated. It also means that the relational symmetrization approaches outlined above are inadequate in the general case. Unfortunately, we are not aware of any work on OSAs of propositional graphical models. However, some existing techniques provide a promising direction. First, basic clustering can group together similar potentials. Second, the low-rank Boolean matrix factorization used for relational approximations can be applied to any graph structure, including graphical models. Third, color passing techniques for exact symmetries operate on propositional models (Kersting et al., 2009, 2014). Combined with early stopping, they can output approximate variable orbits.

[20] Solutions to this problem exist if the soft evidence is on a single unary relation (Bui et al., 2012)

10.6.3 From OSAs to Automorphisms

Given an OSA of our model, we need to compute an automorphism group $\mathfrak{G}$ from it. The obvious choice is to compute the exact automorphisms from the OSA. While this works in principle, it may not be optimal. Let us first consider the following two concepts. When a group $\mathfrak{G}$ operates on a set Ω, only a subset of the elements in Ω can actually be mapped to an element other than itself. When Ω is the set of random variables, we call these elements the *moved variables*. When Ω is the set of potentials in a probabilistic graphical model, we call these the *moved potentials*. It is clear that we want $\mathfrak{G}$ to move many random variables, as this will create the largest jumps and improve the mixing behavior. However, each LMH step comes at a cost: in the second step of the algorithm, the probability of the proposed approximately-symmetric state $\pi(\mathbf{y})$ is estimated. This requires the re-evaluation of all potentials that are moved by $\mathfrak{G}$. Thus, the time complexity of an orbital Metropolis step is linear in the number of moved potentials. It will therefore be beneficial to construct *subgroups* of the automorphism group of the OSA and, in particular, ones that move many variables and few potentials.

10.7 Empirical Evaluation

This section empirically evaluates lifted MCMC, both on symmetric models where we use orbital Markov chains, and on asymmetric models with approximate symmetries, where we use then lifted Metropolis-Hastings algorithm.

10.7.1 Symmetric Model Experiments

We conduct experiments with the well-established social network Markov logic network (the smokes-cancer MLN) exactly as specified in (Singla and Domingos, 2008). Here we created two ground MLNs with 50 and 100, respectively, people in the domain, leading to Markov networks with 2600 and 10200 variables, respectively. Building the ground models took only a fraction of a second. We proceeded to apply the symmetry detection algorithm (Niepert, 2012b) taking 24 and 136 ms, respectively, to compute the irredundant generators of the automorphism group of the models. For n people in the domain, there are $n-1$ irredundant generators of the automorphism group and the group has size $n!$ which is exactly the size of the symmetric group on n. Please note that, based on our observation of indistinguishability of objects on different syntactical levels of the model, it is actually not necessary to use symmetry detection algorithms in this case. The irredundant generators of the symmetric group representing the symmetries on the level of constants can be directly used to compute the irredundant generators for the permutation group representing the symmetries on the level of ground atoms and formulas.

We compared the standard Gibbs sampler, Alchemy's MC-SAT algorithm (Poon and Domingos, 2006), and the orbital Gibbs sampler on the models. The overhead of the product replacement algorithm was again negligible and far outweighed by the faster con-

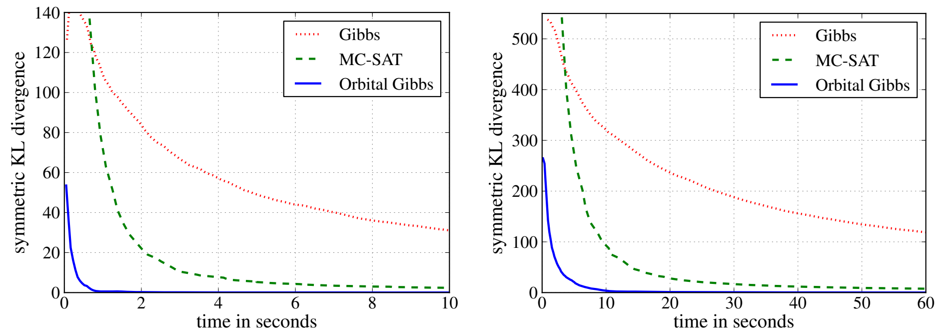


Figure 10.5: The results of the standard Gibbs sampler, Alchemy's MC-SAT algorithm, and the orbital Gibbs sampler for the social network MLN with 50 (top) and 100 (bottom) people in the domain.

vergence of the orbital chain. Figure 10.5 plots the symmetric Kullback-Leibler divergence for the single variable marginals.

Finally, Niepert (2012b) reports additional experiments on using orbital Markov chains to sample independent sets, building on *insert/delete* Markov chains (Luby and Vigoda, 1999; Dyer and Greenhill, 2000).

10.8 Asymmetric Model Experiments

The LMH algorithm is implemented in the GAP algebra system which provides basic algorithms for automorphism groups such as the product replacement algorithm that allows one to sample uniformly from orbits of states (Niepert, 2012b).

For our first experiments, we use the standard WebKB data set, consisting of web pages from four computer science departments (Craven and Slattery, 2001). The data has information about approximately 800 words that appear on 1000 pages, 7 page labels and links between web pages. There are 4 folds, one for each university. We use the standard MLN structure for the WebKB domain, which has MLN formulas of the form shown above, but for all combinations of labels and words, adding up to around 5500 first-order MLN formulas. We learn the MLN parameters using Alchemy.

We consider a collective classification setting, where we are given the link structure and the word content of each web page, and want to predict the page labels. We run Gibbs sampling and the Lifted MCMC algorithm (Niepert, 2012b), and show the average KL divergence between the estimated and true marginals in Figure 10.6. When true marginals are not computable, we used a very long run of a Gibbs sampler for the gold standard marginals. Since every web page contains a unique set of words, the evidence on the word content creates distinct soft evidence on the page labels. Moreover, the link structure is largely asymmetric and, therefore, there are no exploitable exact symmetries and Lifted

MCMC coincides with Gibbs sampling. Next we construct an OSA using a rank-5 approximation of the link structure (Van den Broeck and Darwiche, 2013) and group the potential weights into 6 clusters.

From this OSA we construct a set of automorphisms that is efficient for LMH as follows. First, we compute the exact automorphisms $\mathfrak{G}_1$ of the OSA. Second, we compute the variable orbits of $\mathfrak{G}_1$, grouping together all variables that can be mapped into each other. Then, for every orbit O, we construct a set of automorphisms as follows. We greedily search for a $O' \subseteq O$ such that the symmetric group $\mathfrak{G}_{O'}$ on O' maximizes the ratio between the number of moved variables (i.e., $|O'|$) and the number of moved potentials, while keeping the number of moved potentials bounded by a constant K. This guarantees that $\mathfrak{G}_{O'}$ yields an efficient orbital Metropolis chain. Finally, we remove O' from O and recurse until O is empty. From this set of symmetric groups $\mathfrak{G}_{O'}$, we construct a set of orbital Metropolis chains, each with it own set of moved potentials.

Figure 10.6 shows that the LMH chain, with mixing parameter $\alpha = 4/5$, has a lower KL divergence than Gibbs and Lifted MCMC vs. the number of iterations. Note that there is a slight overhead to LMH because the orbital Metropolis chain is run between base chain steps. Despite this overhead, LMH outperforms the baselines as a function of time. The orbital Metropolis chain accepts approximately 70% of its proposals.

Figure 10.7 illustrates the effect of running Lifted MCMC on OSA, which is the current state-of-the-art approach for asymmetric models. As expected, the drawn sample points produce biased estimates. As the quality of the approximation increases, the bias reduces, but so do the speedups. LMH does not suffer from a bias. Moreover, we observe that its performance is stable across different OSAs (not depicted).

We also ran experiments for two propositional models that are frequently used in real world applications. The first model is a 100x100 ferromagnetic Ising model with constant interaction strength and external field (see Figure 10.1(a) for a 4x4 version). Due to the different potentials induced by the external field, the model has no symmetries. We use the model without external field to compute the approximate symmetries. The automorphism group representing these symmetries is generated by the rotational and reflectional symmetries of the grid model (see Figure 10.1(b)). As in the experiments with the relational models, we used the mixing parameter $\alpha = 4/5$ for the LMH algorithm. Figure 10.8(c) and (d) depicts the plots of the experimental results. The LMH algorithm performs better with respect to the number of iterations and, to a lesser extent, with respect to time.

We also ran experiments on the Chimera model which has recently received some attention as it was used to assess the performance of quantum annealing (Boixo et al., 2013). We used exactly the model as described in Boixo et al. (2013). This model is also asymmetric but can be made symmetric by assuming that all pairwise interactions are identical. The KL divergence vs. number of iterations and vs. time in seconds is plotted in Figure 10.8(a) and (b), respectively. Similar to the results for the Ising model, LMH outperforms Gibbs and

LMCMC both with respect to the number of iterations and wall clock time. In summary, the LMH algorithm outperforms standard sampling algorithms on these propositional models in the absence of any symmetries. We used very simple symmetrization strategies for the experiments. This demonstrates that the LMH framework is powerful and allows one to design state-of-the-art sampling algorithms.

10.9 Conclusions

We have presented a perspective on lifted inference, where instead of directly operating on the space of joint variable assignments, *orbital* Markov chains operate on a symmetry-induced partition of this space. We related lifted MCMC to the notion of lumping of Markov chains. Instead of computing the partition of the state space explicitly which is usually intractable, orbital Markov chains operate on the original state space while having convergence properties identical to the corresponding lumped Markov chain. We want to point out that in the MCMC literature a lifting of a Markov chain (Chen et al., 1999) is *not* the same as what has been coined lifted inference by the statistical relational AI community. Quite the opposite, instead of operating on a more compact state space, lifting in the classical sense introduces additional states. Nevertheless, there might be interesting relationships between lumping, lifting and lifted inference.

We have also presented a Lifted Metropolis-Hastings algorithms capable of mixing two types of Markov chains. The first is a non-lifted base chain, and the second is an orbital Metropolis chain that moves between approximately symmetric states. This allows lifted inference techniques to be applied to asymmetric graphical models.

Numerous extensions to the lifted MCMC framework have been developed in recent years, for example towards exploiting contextual and block-valued symmetries (Anand et al., 2016; Madan et al., 2018), and continuous symmetries (Shariff et al., 2015). Holtzen et al. (2019a) showed how to build an exact lifted inference algorithm that enumerates and counts orbits, and how to sample orbits directly, without the need for Gibbs sampling to move between orbits.

10.10 Acknowledgments

Lifted MCMC for symmetric models first appeared as Niepert (2012b) and Niepert (2012a). Lifted MCMC for asymmetric models first appeared as Van den Broeck and Niepert (2015).

10.11 Appendix: Proof of Theorem 10.3

We first prove (a). Since $\mathcal{M}'$ is aperiodic we have, for each state $x \in \Omega$ and every time step $t \geq 0$, a non-zero probability for the Markov chain $\mathcal{M}'$ to remain in state x at time $t + 1$. At each time $t + 1$, the orbital Markov chain transitions uniformly at random to one of the states in the orbit of the original chain's state at time $t + 1$. Since every state is an element of its own orbit, we have, for every state $x \in \Omega$ and every time step $t \geq 0$, a non-zero probability for the Markov chain $\mathcal{M}$ to remain in state x at time $t+1$. Hence, $\mathcal{M}$ is aperiodic. The proof of statement (b) is accomplished in an analogous fashion and omitted.

Let $P(x, y)$ and $P'(x, y)$ be the probabilities of $\mathcal{M}$ and $\mathcal{M}'$, respectively, to transition from state x to state y. Since π is a reversible distribution for $\mathcal{M}'$ we have that $\pi(x)P'(x, y) = \pi(y)P'(y, x)$ for all states $x, y \in \Omega$. For every state $x \in \Omega$ let $x^{\mathfrak{G}}$ be the orbit of x. Let $\mathfrak{G}_x := \{\mathfrak{g} \in \mathfrak{G} \mid x^{\mathfrak{g}} = x\}$ be the stabilizer subgroup of x with respect to $\mathfrak{G}$. We have that

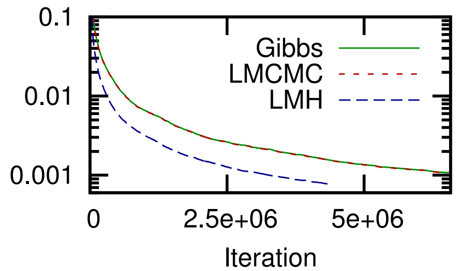
$$\begin{aligned} \sum_{\mathfrak{g}\in\mathfrak{G}} P'(x, y^{\mathfrak{g}}) &= \sum_{y'\in y^{\mathfrak{G}}} |\mathfrak{G}_{y'}| P'(x, y') \\ &= |\mathfrak{G}_y| \sum_{y'\in y^{\mathfrak{G}}} P'(x, y') \\ &= (|\mathfrak{G}|/|y^{\mathfrak{G}}|) \sum_{y'\in y^{\mathfrak{G}}} P'(x, y') \end{aligned} \tag{10.1}$$

where the last two equalities follow from the orbit-stabilizer theorem. We will now prove that $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all states $x, y \in \Omega$. By definition of the orbital Markov chain we have that $\pi(x)P(x, y) = \pi(x)(1/|y^{\mathfrak{G}}|) \sum_{y'\in y^{\mathfrak{G}}} P'(x, y')$ and, by equation (1), we have $\pi(x)(1/|y^{\mathfrak{G}}|) \sum_{y'\in y^{\mathfrak{G}}} P'(x, y')$

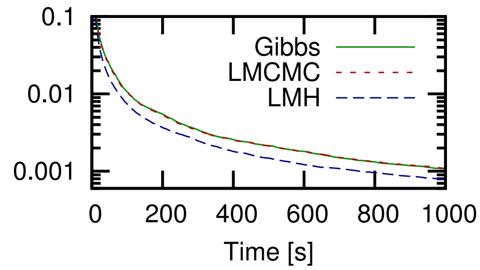
$$\begin{aligned} &= \pi(x)(1/|y^{\mathfrak{G}}|)(|y^{\mathfrak{G}}|/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(x, y^{\mathfrak{g}}) \\ &= \pi(x)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(x, y^{\mathfrak{g}}) \\ &= (1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} \pi(x)P'(x, y^{\mathfrak{g}}). \end{aligned}$$

Since P' is reversible and $\pi(x) = \pi(x^{\mathfrak{g}})$ for all $\mathfrak{g} \in \mathfrak{G}$ we have $(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} \pi(x)P'(x, y^{\mathfrak{g}}) = (1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} \pi(y^{\mathfrak{g}})P'(y^{\mathfrak{g}}, x) = \pi(y)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(y^{\mathfrak{g}}, x)$. Now, since $P'(x, y) = P'(x^{\mathfrak{g}}, y^{\mathfrak{g}})$ for all $x, y \in \Omega$ and all $\mathfrak{g} \in \mathfrak{G}$ by assumption, we have that $\pi(y)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(y^{\mathfrak{g}}, x) = \pi(y)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(y, x^{-\mathfrak{g}}) = \pi(y)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(y, x^{\mathfrak{g}})$ and, again by equation (1), we have $\pi(y)(1/|\mathfrak{G}|) \sum_{\mathfrak{g}\in\mathfrak{G}} P'(y, x^{\mathfrak{g}})$

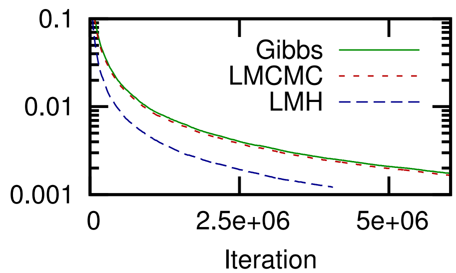
$$\begin{aligned} &= \pi(y)(1/|\mathfrak{G}|)(|\mathfrak{G}|/|x^{\mathfrak{G}}|) \sum_{x'\in x^{\mathfrak{G}}} P'(y, x') \\ &= \pi(y)(1/|x^{\mathfrak{G}}|) \sum_{x'\in x^{\mathfrak{G}}} P'(y, x') = \pi(y)P(y, x). \qquad \square \end{aligned}$$



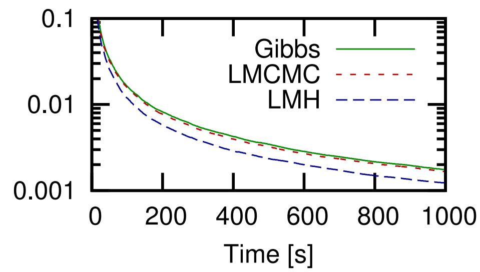
(a) Texas - Iterations



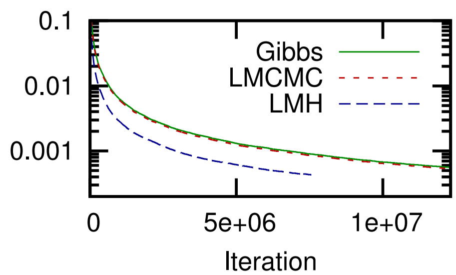
(b) Texas - Time



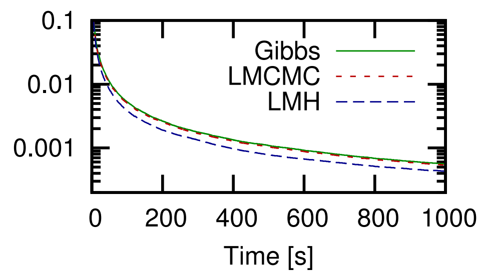
(c) Washington - Iterations



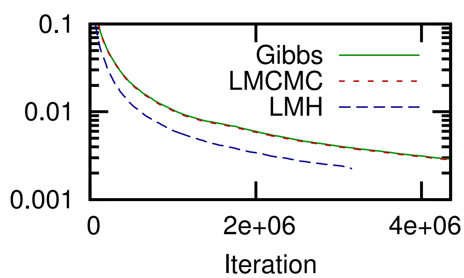
(d) Washington - Time



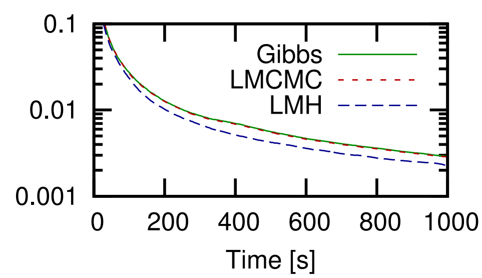
(e) Cornell - Iterations



(f) Cornell - Time



(g) Wisconsin - Iterations



(h) Wisconsin - Time

Figure 10.6: WebKB - KL Divergences

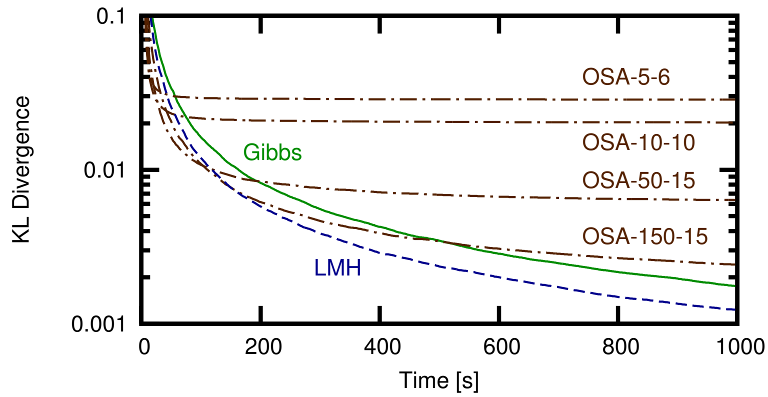
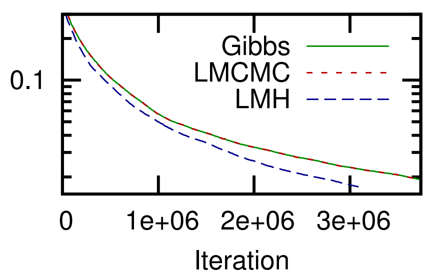
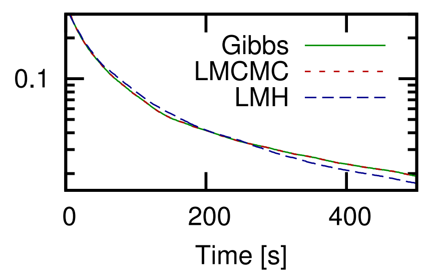


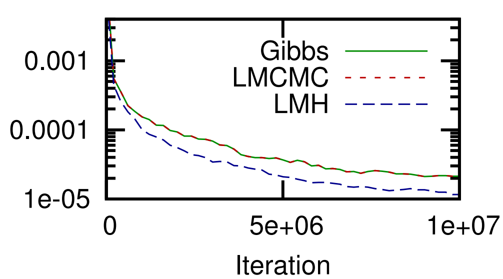
Figure 10.7: LMH vs. over-symmetric approximations (OSA) on WebKB Washington. OSA-*r*-*c* denotes binary evidence of Boolean rank *r* and *c* clusters of formula weights.



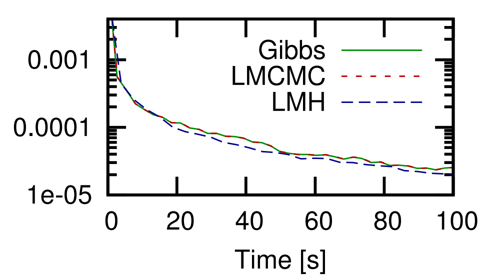
(a) Ising - Iterations



(b) Ising - Time



(c) Chimera - Iterations



(d) Chimera - Time

Figure 10.8: KL Divergences for the propositional models.

Bibliography

Ahmadi, B., Kersting, K., and Hadiji, F. (2010). Lifted belief propagation: Pairwise marginals and beyond. In *Proceedings of the 5th European Workshop on Probabilistic Graphical Models (PGM)*. 232, 275, 341, 343

Ahmadi, B., Kersting, K., and Natarajan, S. (2012). Lifted online training of relational models with stochastic gradient methods. In *ECML PKDD*, 585–600. 89

Ahmadi, B., Kersting, K., and Sanner, S. (2011). Multi-Evidence Lifted Message Passing, with Application to PageRank and the Kalman Filter. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI 2011*. 275, 318

Aji, S. M. and McEliece, R. J. (2001). The generalized distributive law and free energy minimization. In *Proceedings of the 39th Allerton Conference on Communication, Control and Computing*, 672–681. 273

Aloul, F. A., Sakallah, K. A., and Markov, I. L. (2006). Efficient symmetry breaking for boolean satisfiability. *IEEE Transactions on Computers*, 55(5):549–558. 226

Anand, A., Grover, A., Mausam, M., and Singla, P. (2016). Contextual symmetries in probabilistic graphical models. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3560–3568. 198

Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43. 190

Apsel, U. and Brafman, R. I. (2011). Extended lifted inference with joint formulas. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, 11–18. 57

Apt, K. R. and Bezem, M. (1991). Acyclic programs. *New Generation Computing*, 9(3-4):335–363. 6

Arnborg, S., Corneil, D. G., and Proskurowski, A. (1987). Complexity of finding embeddings in a k-tree. *SIAM Journal on Algebraic Discrete Methods*, 8(2):277–284. 99

Atserias, A. and Maneva, E. (2013). Sherali-Adams relaxations and indistinguishability in counting logics. *SIAM Journal of Computation*, 42(1):112–137. 364

Attias, H. (2003). Planning by probabilistic inference. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, AISTATS 2003, Key West, Florida, USA, January 3-6, 2003*. 376, 377

Babai, L. (2016). Graph isomorphism in quasipolynomial time. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC '16)*, 684–697. 355

Babai, L., Erdős, P., and Selkow, S. (1980). Random graph isomorphism. *SIAM Journal on Computing*, 9:628–635. 351, 355

Babai, L. and Kučera, L. (1979). Canonical labelling of graphs in linear average time. In *Annual Symposium on Foundations of Computer Science (FOCS)*, 39–46. 355

Bach, S. H., Broecheler, M., Huang, B., and Getoor, L. (2017). Hinge-loss markov random fields and probabilistic soft logic. *Journal of Machine Learning Research (JMLR)*, 18:1–67. 28, 53

Bahar, R. I., Frohm, E., Gaona, C., Hachtel, G., Macii, E., Pardo, A., and Somenzi, F. (1993). Algebraic Decision Diagrams and their applications. In *IEEE /ACM International Conference on CAD*, 428–432. 393

Bahmani-Oskooee, M. and Brown, F. (2004). Kalman filter approach to estimate the demand for international reserves. *Applied Economics*, 36(15):1655–1668. 327

Balas, E., Ceria, S., and Cornuéjols, G. (1993). A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58:295–324. 364

Battaglia, D., Kolář, M., and Zecchina, R. (2004). Minimizing energy below the glass thresholds. *Physical Review E*, 70:036107. 226

Beame, P., Van den Broeck, G., Gribkoff, E., and Suciu, D. (2015). Symmetric weighted first-order model counting. In *PODS*, 313–328. 104, 110, 155, 156, 158, 159, 164

Belle, V., Passerini, A., and Van den Broeck, G. (2015a). Probabilistic inference in hybrid domains by weighted model integration. In *Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*. 101

Belle, V., Van den Broeck, G., and Passerini, A. (2015b). Hashing-based approximate probabilistic inference in hybrid domains. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence (UAI)*. 101

Bellodi, E., Lamma, E., Riguzzi, F., Santos Costa, V., and Zese, R. (2014). Lifted variable elimination for probabilistic logic programming. *Theory and Practice of Logic Programming (TPLP)*, 14(4–5):681–695. 101

Berkholz, C., Bonsma, P., and Grohe, M. (2013). Tight lower and upper bounds for the complexity of canonical colour refinement. In *Proc. of the 21st Annual European Symposium on Algorithms, Sophia Antipolis, France, September 2-4, 2013.*, 145–156. Springer. 353

Berkholz, C., Bonsma, P., and Grohe, M. (2016). Tight lower and upper bounds for the complexity of canonical colour refinement. *Theory of Computing Systems*, doi:10.1007/s00224-016-9686-0. 355

Berry, A., Blair, J. R., Heggernes, P., and Peyton, B. W. (2004). Maximum cardinality search for computing minimal triangulations of graphs. *Algorithmica*, 39(4):287–298. 99

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236. 43

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007). *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media. 92, 96

Blockeel, H. and De Raedt, L. (1998). Top-down induction of first-order logical decision trees. *Artificial intelligence*, 101(1):285–297. 45

Bodlaender, H. L. (1993). A tourist guide through treewidth. *Acta Cybernetica*, 11(1–2):1–21. 12

Boixo, S., Rønnow, T. F., Isakov, S. V., Wang, Z., Wecker, D., Lidar, D. A., Martinis, J. M., and Troyer, M. (2013). Quantum annealing with more than one hundred qubits. *Nature Physics*, 10(3):218–224. 197

Böker, J. (2019). Color refinement, homomorphisms, and hypergraphs. *ArXiv (CoRR)*, arXiv:1903.12432 [cs.DM]. 352

Bollobás, B. (1982). Distinguishing vertices of random graphs. *Annals of Discrete Mathematics*, 13:33–50. 358

Borgwardt, S., Ceylan, I. I., and Lukasiewicz, T. (2017). Ontology-mediated queries for probabilistic databases. In *Thirty-First AAAI Conference on Artificial Intelligence*. 101

Boutilier, C., Dean, T., and Hanks, S. (1999). Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research (JAIR)*, 11:1–94. 392

Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. (1996). Context-specific independence in Bayesian networks. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, 115–123. 89, 161, 392

Boutilier, C., Reiter, R., and Price, B. (2001). Symbolic dynamic programming for first-order MDPs. In *International Joint Conference on Artificial Intelligence (IJCAI-01)*, 690–697. Seattle. xix, 374, 381, 384, 387, 393

Braun, T. and Möller, R. (2016). Lifted junction tree algorithm. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, 30–42. Springer. 147

Braun, T. and Möller, R. (2017a). Counting and conjunctive queries in the lifted junction tree algorithm. In *International Workshop on Graph Structures for Knowledge Representation and Reasoning*, 54–72. Springer. 147

Braun, T. and Möller, R. (2017b). Preventing groundings and handling evidence in the lifted junction tree algorithm. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, 85–98. Springer. 147

Braunstein, A., Mézard, M., and Zecchina, R. (2005). Survey propagation: An algorithm for satisfiability. *Random Structures and Algorithms*, 27(2):201–226. 217, 220, 221, 222

Braunstein, A. and Zecchina, R. (2004). Survey propagation as local equilibrium equations. *Journal of Statistical Mechanics: Theory and Experiment*, P06007:812–815. 213, 227

Bröcheler, M., Mihalkova, L., and Getoor, L. (2010). Probabilistic similarity logic. In Grünwald, P. and Spirtes, P. (eds.), *Proc. Twenty Sixth Conference on Uncertainty in Artificial Intelligence*, 73–82. AUAI Press. 53

Bruynooghe, M., De Cat, B., Drijkoningen, J., Fierens, D., Goos, J., Gutmann, B., Kimmig, A., Labeeuw, W., Langenaken, S., Landwehr, N., Meert, W., Nuyts, E., Pellegrims, R., Rymenants, R., Segers, S., Thon, I., Van Eyck, J., Van den Broeck, G., Vangansewinkel, T., Van Hove, L., Vennekens, J., Weytjens, T., and De Raedt, L. (2009). An exercise with statistical relational learning systems. In Domingos, P. and Kersting, K. (eds.), *Proceedings of the International Workshop on Statistical Relational Learning (SRL-09)*, 1–3. 16

Buchholz, P. (1994). Exact and ordinary lumpability in finite Markov chains. *Journal of Applied Probability*, 31(1):59–75. 187

Buchman, D. and Poole, D. (2015). Representing aggregators in relational probabilistic models. In *Proc. Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*. 10, 29

Buchman, D. and Poole, D. (2016). Negation without negation in probabilistic logic programming. In *Proc. 15th International Conference on Principles of Knowledge Representation and Reasoning*. 31

Bui, H., Huynh, T., and Riedel, S. (2013a). Automorphism groups of graphical models and lifted variational inference. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*. 172, 176

Bui, H., Huynh, T., and Sontag, D. (2014). Lifted tree-reweighted variational inference. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*. 304

Bui, H. B., Huynh, T. N., and de Salvo Braz, R. (2012). Exact lifted inference with distinct soft evidence on every object. In *AAAI*. 164, 170, 174, 184, 185, 193, 194, 239, 315

Bui, H. H., Huynh, T. N., and Riedel, S. (2013b). Automorphism groups of graphical models and lifted variational inference. In *UAI*, 132. 89, 164

Buntine, W. L. (1994). Operations for learning with graphical models. *J. Artif. Intell. Res. (JAIR)*, 2:159–225. 8, 53, 115

Burgers, G., van Leeuwen, P. J., Evensen, G., Burgers, G., and Burgers, G. (1998). On the analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126:1719–1724. 327

Cai, J.-Y., Fürer, M., and Immerman, N. (1992). An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12:389–410. 352

Carbonetto, P., Kisyński, J., Chiang, M., and Poole, D. (2009). Learning a contingently acyclic, probabilistic relational model of a social network. TR-2009-08, Univ of British Columbia, Dept of Comp Sci. 126

Carbonetto, P., Kisynski, J., de Freitas, N., and Poole, D. (2005). Nonparametric Bayesian logic. In *Proc. 21st Conference on Uncertainty in AI (UAI)*. 37

Cardon, A. and Crochemore, M. (1982). Partitioning a graph in $O(|A| \log_2 |V|)$. *Theoretical Computer Science*, 19(1):85 – 98. 350, 353

Celler, F., Leedham-Green, C. R., Murray, S. H., Niemeyer, A. C., and O'brien, E. (1995). Generating random elements of a finite group. *Communications in Algebra*, 23(13):4931–4948. 185, 188, 189

Ceylan, I. I., Darwiche, A., and Van den Broeck, G. (2016). Open-world probabilistic databases. In *Description Logics*. 101, 106

Chang, C. and Keisler, J. (1990). *Model Theory*. Elsevier, Amsterdam, Holland. 381

Chang, C. L. and Lee, R. C. T. (1973). *Symbolic Logic and Mechanical Theorem Proving*. Academic Press. 13

Chavira, M. and Darwiche, A. (2008). On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6):772–779. 11, 95

Chavira, M., Darwiche, A., and Jaeger, M. (2006). Compiling relational Bayesian networks for exact inference. *International Journal of Approximate Reasoning*, 42(1-2):4–20. 95

Chen, F., Lovász, L., and Pak, I. (1999). Lifting Markov chains to speed up mixing. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, STOC '99, 275–281. 198

Chen, Y., Ruozzi, N., and Natarajan, S. (2019). Lifted message passing for hybrid probabilistic inference. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 5701–5707. International Joint Conferences on Artificial Intelligence Organization. 343

Chen, Y., Yang, Y., Natarajan, S., and Ruozzi, N. (2020). Hybrid lifted variational inference. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-20*. 344

Cheng, Q., Liu, Q., Chen, F., and Ihler, A. T. (2013). Variational planning for graph-based MDPs. In *Proc. Advances in Neural Information Processing Systems*, 2976–2984. 376, 377

Chieu, H. and Lee, W. (2009). Relaxed survey propagation for the weighted maximum satisfiability problem. *Journal of Artificial Intelligence Research (JAIR)*, 36:229–266. 226

Choi, A., Chavira, M., and Darwiche, A. (2007). Node splitting: A scheme for generating upper bounds in Bayesian networks. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 57–66. 273

Choi, A. and Darwiche, A. (2006). An edge deletion semantics for belief propagation and its practical impact on approximation quality. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence,*, 1107–1114. 261, 262, 273

Choi, A. and Darwiche, A. (2008). Approximating the partition function by deleting and then correcting for model edges. In *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence (UAI)*, 79–87. 273

Choi, A. and Darwiche, A. (2011). Relax, compensate and then recover. In Onada, T., Bekki, D., and McCready, E. (eds.), *New Frontiers in Artificial Intelligence*, volume 6797 of *Lecture Notes in Computer Science*, 167–180. Springer Berlin / Heidelberg. 259, 263, 273

Choi, J. and Amir, E. (2012). Lifted relational variational inference. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 196–206. 256, 257, 318

Choi, J., Amir, E., Xu, T., and Valocchi, A. J. (2015). Learning relational Kalman filtering. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2539–2546. 318

Choi, J., de Salvo Braz, R., and Bui, H. H. (2011a). Efficient methods for lifted inference with aggregate factors. In *AAAI*. 89

Choi, J., Guzman-Rivera, A., and Amir, E. (2011b). Lifted relational kalman filtering. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence(IJCAI)*, 2092–2099. 318

Choi, J., Hill, D., and Amir, E. (2010). Lifted inference for relational continuous models. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*, 126–134. 57, 318, 327, 331, 336

Clark, K. L. (1978). Negation as failure. In Gallaire, H. and Minker, J. (eds.), *Logic and Databases*, 293–322. Springer. 6, 30, 142

Clautiaux, F., Moukrim, A., Nègre, S., and Carlier, J. (2004). Heuristic and metaheuristic methods for computing graph treewidth. *RAIRO-Operations Research*, 38(1):13–26. 99

Cozman, F. G. (2004). Axiomatizing noisy-or. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, 979–980. 143

Craven, M. and Slattery, S. (2001). Relational learning with statistical predicate invention: Better models for hypertext. *Machine Learning Journal*, 43(1/2):97–119. 196

Crawford, J. (1992). A theoretical analysis of reasoning by symmetry in first-order logic. In *Proceedings of the Workshop on Tractable Reasoning*. 226

Crawford, J. M., Ginsberg, M. L., Luks, E. M., and Roy, A. (1996). Symmetry-breaking predicates for search problems. In *Proceedings of the 5th Conference on Principles of Knowledge Representation and Reasoning*, 148–159. 221

Cui, H., Keller, T., and Khardon, R. (2019). Stochastic planning with lifted symbolic trajectory optimization. In *ICAPS*. 394

Cui, H., Marinescu, R., and Khardon, R. (2018). From stochastic planning to marginal MAP. In *NIPS*, 3085–3095. 376, 394

Dalvi, N. and Suciu, D. (2012). The dichotomy of probabilistic inference for unions of conjunctive queries. *Journal of the ACM (JACM)*, 59(6):30. 106

Dalvi, N. N. and Suciu, D. (2004). Efficient query evaluation on probabilistic databases. In Nascimento, M. A., Özsu, M. T., Kossmann, D., Miller, R. J., Blakeley, J. A., and Schiefer, K. B. (eds.), *Proceedings of the 30th International Conference on Very Large Databases (VLDB-04)*, 864–875. Morgan Kaufmann Publishers. 21

Damien, P. and Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2):206–215. 133

Dantsin, E. (1991). Probabilistic logic programs and their semantics. In Voronkov, A. (ed.), *Proceedings of the First Russian Conference on Logic Programming*, volume 592 of *Lecture Notes in Computer Science*, 152–164. Springer. 21

Darwiche, A. (2001). Recursive conditioning. *Artificial Intelligence*, 126(1-2):5–41. 12, 89

Darwiche, A. (2003). A differential approach to inference in Bayesian networks. *Journal of the ACM (JACM)*, 50(3):280–305. 147, 149

Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press. 99, 161, 165

Darwiche, A. and Marquis, P. (2002). A knowledge compilation map. *J. Artif. Intell. Res. (JAIR)*, 17:229–264. 147, 148, 149

Davis, J. and Domingos, P. (2009). Deep transfer via second-order Markov logic. In *Proc. International Conference on Machine Learning (ICML)*. ACM Press, Montréal, Canada. 48

Davis, J. and Domingos, P. (2011). Deep transfer: A Markov logic approach. *AI Magazine*, 32(1):51–52. 48

Davis, M., Logemann, G., and Loveland, D. (1962). A machine program for theorem proving. *Communications of the ACM*, 5(7):394–397. 12

De Raedt, L., Frasconi, P., Kersting, K., and Muggleton, S. (eds.) (2008). *Probabilistic Inductive Logic Programming — Theory and Applications*, volume 4911 of *Lecture Notes in Artificial Intelligence*. Springer. 57, 174

De Raedt, L., Kersting, K., Natarajan, S., and Poole, D. (2016). *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. Morgan & Claypool. 3, 16, 28, 39

De Raedt, L. and Kimmig, A. (2015). Probabilistic (logic) programming concepts. *Machine Learning*, 100(1):5–47. 16, 17, 21, 24, 29

De Raedt, L., Kimmig, A., and Toivonen, H. (2007). ProbLog: A probabilistic Prolog and its application in link discovery. In Veloso, M. M. (ed.), *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI-07)*, 2462–2467. Morgan Kaufmann Publishers. 3, 16, 17, 21, 24, 40, 53, 101, 137, 141

de Salvo Braz, R., Amir, E., and Roth, D. (2005). Lifted first-order probabilistic inference. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, 1319–1325. 57, 89, 105, 107, 174, 270, 275, 276, 318, 319, 320, 327, 335

de Salvo Braz, R., Amir, E., and Roth, D. (2006). MPE and partial inversion in lifted probabilistic variable elimination. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, 1123–1130. 106, 319

de Salvo Braz, R., Amir, E., and Roth, D. (2007). Lifted first-order probabilistic inference. In Getoor, L. and Taskar, B. (eds.), *An Introduction to Statistical Relational Learning*, 433–451. MIT Press. 66, 71, 72, 79, 85, 113, 118, 120, 125

de Salvo Braz, R., Natarajan, S., Bui, H., Shavlik, J., and Russell, S. (2009). Anytime lifted belief propagation. *Proc. SRL-09*. 275

de Salvo Braz, R., O'Reilly, C., Gogate, V., and Dechter, R. (2016). Probabilistic inference modulo theories. *arXiv preprint arXiv:1605.08367*. 101

Dean, T. and Kanazawa, K. (1989). A Model for Reasoning about Persistence and Causation. *Computational Intelligence*, 5:142–150. 392

Dearden, R. and Boutilier, C. (1997). Abstraction and approximate decision theoretic planning. *Artificial Intelligence*, 89(1-2):219–283. 392

Dechter, R. (1996). Bucket elimination: A unifying framework for probabilistic inference. In Horvitz, E. and Jensen, F. (eds.), *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96)*, 211–219. 12

Dechter, R. (2003). *Constraint Processing*. Morgan Kaufmann. 12, 67, 99

Dechter, R., Kask, K., and Mateescu, R. (2002). Iterative join-graph propagation. In *Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence (UAI)*, 128–136. 273

Dechter, R. and Mateescu, R. (2007). And/or search spaces for graphical models. *Artif. Intell.*, 171(2-3):73–106. 149

Dechter, R. and Rish, I. (2003). Mini-buckets: A general scheme for bounded inference. *Journal of the ACM (JACM)*, 50(2):107–153. 273

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society B*, 39:1–38. 238

Derisavi, S., Hermanns, H., and Sanders, W. H. (2003). Optimal state-space lumping in Markov chains. *Inf. Process. Lett.*, 87(6):309–315. 187, 188

Diaconis, P. and Freedman, D. (1980). De Finetti's generalizations of exchangeability. In *Studies in Inductive Logic and Probability*, volume II. 162, 165

Diez, F. J. (1993). Parameter adjustment in Bayes networks. the generalized noisy or-gate. In *Proc. ninth UAI*, 99–105. 117

Díez, F. J. and Galán, S. F. (2003). Efficient computation for the noisy max. *International Journal of Intelligent Systems*, 18(2):165–177. 32, 121, 125, 129, 130, 132, 143

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books, New York:. 11

Domingos, P., Kok, S., Lowd, D., Poon, H., Richardson, M., and Singla, P. (2008). Markov logic. In Raedt, L. D., Frasconi, P., Kersting, K., and Muggleton, S. (eds.), *Probabilistic Inductive Logic Programming*, 92–117. Springer. 93

Domingos, P. and Lowd, D. (2009). *Markov Logic: An Interface Layer for Artificial Intelligence*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool. 3, 10, 260

Domshlak, C. and Hoffmann, J. (2006). Fast probabilistic planning through weighted model counting. In *Proceedings of the International Conference on Automated Planning and Scheduling*. 376

Dos Martires, P. Z., Dries, A., and De Raedt, L. (2019). Exact and approximate weighted model integration with probability density functions using knowledge compilation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7825–7833. 101

Doshi, F. and Roy, N. (2008). The permutable POMDP: Fast solutions to POMDPs for preference elicitation. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*. Estoril, Portugal. 394

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001). *Sequential Monte Carlo in Practice*. Springer-Verlag. 13

Driessens, K. and Dzeroski, S. (2002). Integrating experimentation and guidance in relational reinforcement learning. In *International Conference on Machine Learning (ICML)*, 115–122. 394

Duchi, J., Tarlow, D., Elidan, G., and Koller, D. (2007). Using combinatorial optimization within max-product belief propagation. In *Advances in Neural Information Processing Systems 19 (NIPS)*, 369–376. 227

Duris, D. (2012). Some characterizations of γ and β-acyclicity of hypergraphs. *Information Processing Letters*, 112(16):617–620. 156

Dyer, M. and Greenhill, C. (2000). On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49. 196

Dzeroski, S., DeRaedt, L., and Driessens, K. (2001). Relational reinforcement learning. *Machine Learning Journal (MLJ)*, 43:7–52. 394

Eaton, F. and Ghahramani, Z. (2009). Choosing a variable to clamp. In *International Conference on Artificial Intelligence and Statistics*, 145–152. 237

Elidan, G. and Globerson, A. (2010). Summary of the 2010 UAI approximate inference challenge. `http://www.cs.huji.ac.il/project/UAI10/`. 259, 273

Erdős, P. and Rényi, A. (1963). Asymmetric graphs. *Acta Mathematica Hungarica*, 14(3):295–315. 194

Esseen, C.-G. (1942). On the liapunoff limit of error in the theory of probability. *Arkiv foer Matematik, Astronomi, och Fysik*, A28(9):1–19. 135

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162. 327

Fages, F. (1994). Consistency of Clark's completion and existence of stable models. *Journal of Methods of Logic in Computer Science*, 1:51–60. 142

Fagin, R. (1983). Degrees of acyclicity for hypergraphs and relational database schemes. *Journal of the ACM (JACM)*, 30(3):514–550. 158

Feynman, R. P. (1987). Negative probability. *Quantum implications: essays in honour of David Bohm*, 235–248. 143

Fierens, D., Van den Broeck, G., Renkens, J., Shterionov, D., Gutmann, B., Thon, I., Janssens, G., and De Raedt, L. (2015). Inference and learning in probabilistic logic programs using weighted Boolean formulas. *Theory and Practice of Logic Programming*, 15(03):358–401. 21, 24, 101, 141

Fierens, D., Van den Broeck, G., Thon, I., Gutmann, B., and De Raedt, L. (2011). Inference in probabilistic logic programs using weighted CNFs. In Cozman, F. G. and Pfeffer, A. (eds.), *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI-11)*, 211–220. AUAI Press. 95

Fikes, R. E. and Nilsson, N. J. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *AI Journal*, 2:189–208. 374

Freedman, R., de Salvo Braz, R., Bui, H., and Natarajan, S. (2012). Initial empirical evaluation of anytime lifted belief propagation. *Proceedings of Statistical Relational AI (StaRAI) workshop*. 275

Friedman, N., Getoor, L., Koller, D., and Pfeffer, A. (1999). Learning probabilistic relational models. In *Proc. of the Sixteenth International Joint Conference on Artificial Intelligence*, 1300–1307. Sweden: Morgan Kaufmann. 39, 319, 327

Friedman, T. and Van den Broeck, G. (2018). Approximate knowledge compilation by online collapsed importance sampling. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*. 95

Friedman, T. and Van den Broeck, G. (2019). On constrained open-world probabilistic databases. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*. 101, 106

Friedman, T. and Van den Broeck, G. (2020). Symbolic querying of vector spaces: Probabilistic databases meets relational embeddings. In *Ninth International Workshop on Statistical Relational AI (StarAI)*. 101

Fuhr, N. (2000). Probabilistic Datalog: Implementing logical information retrieval for advanced applications. *Journal of the American Society for Information Science (JASIS)*, 51(2):95–110. 21

Furmston, T. and Barber, D. (2010). Variational methods for reinforcement learning. In *Proceedings of the International Conference on Artificial Intelligence and Statistics, AISTATS*, 241–248. 376, 377

GAP (2008). *GAP – Groups, Algorithms, and Programming, Version 4.4.12*. The GAP Group. 181, 189

Gartner, T., Driessens, K., and Ramon, J. (2006). Graph kernels and Gaussian processes for relational reinforcement learning. *Machine Learning Journal (MLJ)*, 64:91–119. 394

Gartner, T., Flach, P., and Wrobel, S. (2003). On graph kernels: Hardness results and efficient alternatives. In Scholkopf, I. M. W. B. (ed.), *Proceedings of the 16th Annual Conference on Computational Learning Theory and the 7th Kernel Workshop*, 129–143. 368

Gehrke, M., Braun, T., and Möller, R. (2018). Lifted dynamic junction tree algorithm. In *International Conference on Conceptual Structures*, 55–69. Springer. 147

Getoor, L., Friedman, N., Koller, D., and Pfeffer, A. (2001). Learning probabilistic relational models. *Relational Data Mining, S. Dzeroski and N. Lavrac, Eds.* 39, 40, 41

Getoor, L. and Taskar, B. (eds.) (2007). *An Introduction to Statistical Relational Learning*. MIT Press. 14, 39, 57, 174, 319

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computer Sciences and Statistics Proceedings the 23rd Symposium on the Interface between*, 571–578. 133

Gkirtzou, K., Honorio, J., Samaras, D., Goldstein, R. Z., and Blaschko, M. B. (2013). fMRI analysis with sparse weisfeiler-lehman graph statistics. In *MLMI*, volume 8184 of *Lecture Notes in Computer Science*, 90–97. Springer. 370

Globerson, A. and Jaakkola, T. (2007). Fixing max-product: Convergent message passing algorithms for map LP-relaxations. In *Proc. of the 21st Annual Conf. on Neural Inf. Processing Systems (NIPS)*. 238, 305

Gogate, V. and Domingos, P. (2010). Exploiting logical structure in lifted probabilistic inference. In *Proceedings of the first international workshop on statistical relational AI (StarAI)*, 1–8. 89, 105, 140, 148

Gogate, V. and Domingos, P. (2011). Probabilistic theorem proving. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, 256–265. 101

Gogate, V., Jha, A. K., and Venugopal, D. (2012). Advances in lifted importance sampling. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, 8–14. 174, 185

Goldberg, L. A. (2001). Computation in permutation groups: counting and randomly sampling orbits. In *Surveys in Combinatorics*, 109–143. Cambridge University Press. 188

Gomes, C., Hoffmann, J., Sabharwal, A., and Selman, B. (2007). From sampling to model counting. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, 2293–2299. 224, 234

Gomes, T. and Costa, V. S. (2012). Evaluating inference algorithms for the prolog factor language. In *International Conference on Inductive Logic Programming*, 74–85. Springer. 101

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., and Tenenbaum, J. B. (2008). Church: A language for generative models. In McAllester, D. A. and Myllymäki, P. (eds.), *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-08)*, 220–229. AUAI Press. 36, 53

Gotze, F. (1991). On the rate of convergence in the multivariate clt. *The Annals of Probability*, 19(2):724–739. 135

Gretton, C. and Thiebaux, S. (2004). Exploiting first-order regression in inductive policy selection. In *Uncertainty in Artificial Intelligence (UAI-04)*, 217–225. Banff, Canada. 394

Gribkoff, E., Suciu, D., and Van den Broeck, G. (2014a). Lifted probabilistic inference: A guide for the database researcher. *Bulletin of the Technical Committee on Data Engineering*, 37(3):6–17. 101

Gribkoff, E., Van den Broeck, G., and Suciu, D. (2014b). Understanding the complexity of lifted inference and asymmetric weighted model counting. In *Conference on Uncertainty in Artificial Intelligence (UAI)*. 101

Grohe, M. (1998). Fixed-point logics on planar graphs. In *Proceedings of the 13th IEEE Symposium on Logic in Computer Science*, 6–15. 358

Grohe, M. (2017). *Descriptive Complexity, Canonisation, and Definable Graph Structure Theory*, volume 47 of *Lecture Notes in Logic*. Cambridge University Press. 352

Grohe, M., Kersting, K., Mladenov, M., and Selman, E. (2014). Dimension reduction via colour refinement. In Schulz, A. and Wagner, D. (eds.), *Proceedings of the 22nd Annual European Symposium on Algorithms*, volume 8737, 505–516. 365, 366, 367, 368

Grohe, M. and Lindner, P. (2019). Probabilistic databases with an infinite open-world assumption. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, 17–31. 101

Grohe, M. and Otto, M. (2015). Pebble games and linear equations. *Journal of Symbolic Logic*, 80(3):797–844. 364

Gupta, R., Diwan, A. A., and Sarawagi, S. (2007). Efficient inference with cardinality-based clique potentials. In *Proc. 24th ICML*, 329–336. 118, 174

Gutmann, B. and Kersting, K. (2006). Tildecrf: Conditional random fields for logical sequences. In *ECML*. 40, 45

Hadiji, F., Ahmadi, B., and Kersting, K. (2011). Efficient sequential clamping for lifted message passing. In *Proceedings of the 34th Annual German Conference on AI (KI)*. 275

Häggström, O. (2002). *Finite Markov chains and algorithmic applications*. London Mathematical Society student texts. Cambridge University Press. 190, 191

Halpern, J. Y. (2003). *Reasoning about Uncertainty*. MIT Press. 90

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceeding of the 30th Annual Conference on Neural Information Processing Systems*, 1025–1035. 370

Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices. *Unpublished*. 7

Hazan, T. and Shashua, A. (2008). Convergent message-passing algorithms for inference over general graphs with convex free energies. In *Proceedings of the Twenty-Forth Conference in Uncertainty in Artificial Intelligence, (UAI)*, 264–273. 303

Heckerman, D., Chickering, D., Meek, C., Rounthwaite, R., and Kadie, C. (2000). Dependency networks for inference, collaborative filtering, and data visualization. *J. Mach. Learn. Res.*, 1:49–75. 12

Heckerman, D., Meek, C., and Koller, D. (2004). Probabilistic models for relational data. Technical Report MSR-TR-2004-30, Microsoft Research. 40

Hescott, B. and Khardon, R. (2015). The complexity of reasoning with FODD and GFODD. *Artificial Intelligence*. 393

Heskes, T. (2006). Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies. *Journal of Artificial Intelligence Research (JAIR)*, 26:153–190. 281

Hill, D. J., Minsker, B. S., and Amir, E. (2009). Real-time Bayesian anomaly detection in streaming environmental data. *Water Resources Research*, 45:W00D28. 318

Hoey, J., St-Aubin, R., Hu, A., and Boutilier, C. (1999). SPUDD: Stochastic planning using decision diagrams. In *Uncertainty in Artificial Intelligence (UAI-99)*, 279–288. Stockholm. 393

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *Ann. Statist.*, 36(3):1171–1220. 368

Hölldobler, S., Karabaev, E., and Skvortsova, O. (2006). FluCaP: A heuristic search planner for first-order MDPs. *Journal of Artificial Intelligence Research (JAIR)*, 27:419–439. 393

Holtzen, S., Millstein, T., and Van den Broeck, G. (2019a). Generating and sampling orbits for lifted probabilistic inference. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*. 198

Holtzen, S., Millstein, T., and Van den Broeck, G. (2019b). Symbolic exact inference for discrete probabilistic programs. *arXiv.cs.PL*, 1904.02079. 95

Holtzen, S., Van den Broeck, G., and Millstein, T. (2018). Sound abstraction and decomposition of probabilistic programs. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 95

Hopcroft, J. (1971). An n log n algorithm for minimizing states in a finite automaton. In Kohavi, Z. and Paz, A. (eds.), *Theory of Machines and Computations*, 189–196. Academic Press. 353

Howard, R. A. (1988). Decision analysis: Practice and promise. *Management Science*, 34(6):679–695. 6

Huang, J. and Darwiche, A. (2007). The language of search. *Journal of Artificial Intelligence Research (JAIR)*, 29:191–219. 148

Huth, M. and Ryan, M. (2004). *Logic in Computer Science: Modelling and Reasoning About Systems*. Cambridge University Press. 30

Huynh, T. N. and Mooney, R. J. (2008). Discriminative structure and parameter learning for Markov logic networks. In *Proc. of the international conference on machine learning*. 44

Ihler, A., Fisher III, J., and Willsky, A. (2005). Loopy belief propagation: Convergence and effects of message errors. *Journal of Machine Learning Research (JMLR)*, 6:905–936. 208, 242, 243

Immerman, N. and Lander, E. (1990). Describing graphs: A first-order approach to graph canonization. In Selman, A. (ed.), *Complexity theory retrospective*, 59–81. Springer-Verlag. 352, 355, 358, 360

Issakkimuthu, M., Fern, A., Khardon, R., Tadepalli, P., and Xue, S. (2015). Hindsight optimization for probabilistic planning with factored actions. In *ICAPS*. 376

Jaeger, M. (1997). Relational Bayesian networks. In Geiger, D. and Shenoy, P. P. (eds.), *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 266–273. Morgan Kaufmann Publishers. 33, 39

Jaeger, M. (2002). Relational Bayesian networks: a survey. *Electronic Articles in Computer and Information Science*, 6. 117

Jaeger, M. (2015). Lower complexity bounds for lifted inference. *Theory and Practice of Logic Programming*, 15(2):246–263. 158, 159

Jaeger, M. and Van den Broeck, G. (2012). Liftability of probabilistic inference: Upper and lower bounds. In *Proceedings of the 2nd International Workshop on Statistical Relational AI (StaRAI)*, 55–62. 143, 158, 159, 164, 173, 174

Jaimovich, A., Meshi, O., and Friedman, N. (2007). Template-based Inference in Symmetric Relational Markov Random Fields. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 191–199. 274, 304

Jain, A., Friedman, T., Kuzelka, O., Van den Broeck, G., and De Raedt, L. (2019). Scalable rule learning in probabilistic knowledge bases. In *The 1st Conference On Automated Knowledge Base Construction (AKBC)*. 101

Janhunen, T. (2004). Representing normal programs with clauses. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, volume 16, 358. 142

Jensen, F. V., Lauritzen, S. L., and Olesen, K. G. (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational statistics quarterly 4*, 269–282. 147

Jernite, Y., Rush, A. M., and Sontag, D. (2015). A fast variational approach for learning Markov random field language models. In *Proc. International Conference on Machine Learning (ICML)*. 89

Jha, A., Gogate, V., Meliou, A., and Suciu, D. (2010). Lifted inference seen from the other side : The tractable features. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS)*, 973–981. 89, 105, 163, 174

Jha, A. and Suciu, D. (2012). Probabilistic databases with MarkoViews. *Proceedings of the VLDB Endowment*, 5(11):1160–1171. 143

Jordan, M. I. (2010). Bayesian nonparametric learning: Expressive priors for intelligent systems. In Dechter, R., Geffner, H., and Halpern, J. Y. (eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, 167–186. College Publications. 53

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1997). An introduction to variational methods for graphical models. Technical report, MIT Computational Cognitive Science. 13

Joshi, S., Kersting, K., and Khardon, R. (2010). Self-taught decision theoretic planning with first order decision diagrams. In *Proc. of ICAPS*, 89–96. 393

Joshi, S., Kersting, K., and Khardon, R. (2011). Decision theoretic planning with generalized first order decision diagrams. *AIJ*, 175:2198–2222. 381, 383, 385, 387, 390, 393

Joshi, S. and Khardon, R. (2008). Stochastic planning with first order decision diagrams. In *Proc. of ICAPS*, 156–163. 393

Joshi, S., Khardon, R., Raghavan, A., Tadepalli, P., and Fern, A. (2013). Solving relational MDPs with exogenous events and additive rewards. In *ECML*. 381, 383, 386, 390, 393

Joshi, S., Schermerhorn, P. W., Khardon, R., and Scheutz, M. (2012). Abstract planning for reactive robots. In *ICRA*, 4379–4384. 394

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45. 327

Kang, B. K. and Kim, K. (2012). Exploiting symmetries for single- and multi-agent partially observable stochastic domains. *Artif. Intell.*, 182-183:32–57. 394

Karabaev, E. and Skvortsova, O. (2005). A heuristic search algorithm for solving first-order MDPs. In *Uncertainty in Artificial Intelligence (UAI-05)*, 292–299. Edinburgh, Scotland. 393

Karp, R. (1972). Reducibilities among combinatorial problems. In Miller, R. and Thatcher, J. (eds.), *Complexity of Computer Computations*, 85–103. Plenum Press, New York. 355

Kask, K. and Dechter, R. (2001). A general scheme for automatic generation of search heuristics from specification dependencies. *Artificial Intelligence*, 129(1-2):91–131. 273

Kazemi, S. M., Buchman, D., Kersting, K., Natarajan, S., and Poole, D. (2014). Relational logistic regression. In *Proc. 14th International Conference on Principles of Knowledge Representation and Reasoning (KR-2014)*. 10, 31, 32, 33

Kazemi, S. M., Kimmig, A., Van den Broeck, G., and Poole, D. (2016). New liftable classes for first-order probabilistic inference. In *Advances in Neural Information Processing Systems*, 3117–3125. xvii, 89, 109, 147, 149, 152, 153, 155, 156, 157, 158, 164, 392

Kazemi, S. M., Kimmig, A., Van den Broeck, G., and Poole, D. (2017). Domain recursion for lifted inference with existential quantifiers. In *Workshop on Statistical Relational Artificial Intelligence (StaRAI)*. 110, 155, 157, 158, 159, 392

Kazemi, S. M. and Poole, D. (2014). Elimination ordering in first-order probabilistic inference. In *AAAI*. 89, 110

Kersting, K., Ahmadi, B., and Natarajan, S. (2009). Counting belief propagation. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, 277–284. 89, 185, 194, 206, 211, 274, 277, 304, 338, 339, 343

Kersting, K. and De Raedt, L. (2001). Bayesian logic programs. *CoRR*, cs.AI/0111058. 40, 41, 51

Kersting, K. and De Raedt, L. (2008). Basic principles of learning Bayesian logic programs. In *Probabilistic Inductive Logic Programming - Theory and Applications*, 189–221. Springer-Verlag. 45

Kersting, K., De Raedt, L., and Raiko, T. (2006). Logical hidden Markov models. *Journal of Artificial Intelligence Research (JAIR)*, 25:425–456. 31, 33, 36, 331

Kersting, K. and Driessens, K. (2008). Non-parametric policy gradients: A unified treatment of propositional and relational domains. In *Proc. International Conference on Machine Learning (ICML)*. 394

Kersting, K., El Massaoudi, Y., Hadiji, F., and Ahmadi, B. (2010). Informed lifting for message-passing. In *AAAI*. 248, 275

Kersting, K., Mladenov, M., Garnett, R., and Grohe, M. (2014). Power iterated color refinement. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 1904–1910. 193, 194

Kersting, K., Mladenov, M., and Tokmakov, P. (2015). Relational linear programs. *Artificial Intelligence Journal*. 256

Kersting, K., van Otterlo, M., and de Raedt, L. (2004). Bellman goes relational. In *International Conference on Machine Learning (ICML-04)*, 465–472. ACM Press. 393

Khardon, R. (1999a). Learning action strategies for planning domains. *Artificial Intelligence*, 113(1-2):125–148. 394

Khardon, R. (1999b). Learning to take actions. *Machine Learning*, 35:57–90. 394

Khosravi, H., Schulte, O., Hu, J., and Gao, T. (2012). Learning compact Markov logic networks with decision trees. *Machine Learning*, 89(3):257–277. 47

Khot, T., Natarajan, S., Kersting, K., and Shavlik, J. (2011). Learning Markov logic networks via functional gradient boosting. In *ICDM*. 47

Khot, T., Natarajan, S., Kersting, K., and Shavlik, J. (2015). Gradient-based boosting for statistical relational learning: the Markov logic network and missing data cases. *Machine Learning*, 100(1):75–100. 47

Kiefer, S., Ponomarenko, I., and Schweitzer, P. (2017). The Weisfeiler-Leman dimension of planar graphs is at most 3. In *Proceedings of the 32nd ACM-IEEE Symposium on Logic in Computer Science*. 358

Kimmig, A., Bach, S., Broecheler, M., Huang, B., and Getoor, L. (2012). A short introduction to probabilistic soft logic. In *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*, 1–4. 28

Kimmig, A., Mihalkova, L., and Getoor, L. (2015). Lifted graphical models: A survey. *Machine Learning*, 99(1):1–45. 16, 17

Kimmig, A., Van den Broeck, G., and De Raedt, L. (2011). An algebraic Prolog for reasoning about possible worlds. In Burgard, W. and Roth, D. (eds.), *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-11)*, 209–214. AAAI Press. 21

Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*. 370

Kisa, D., Van den Broeck, G., Choi, A., and Darwiche, A. (2014). Probabilistic sentential decision diagrams. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 1–10. 95, 147

Kisynski, J. and Poole, D. (2009). Constraint processing in lifted probabilistic inference. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, 293–302. 58, 68, 89, 101, 108

Kjaerulff, U. (1990). Triangulation of graphs–algorithms giving small total state space. Technical Report Research Report R-90-09, Aalborg University. 99

Kok, S. and Domingos, P. (2005). Learning structure of Markov logic networks. In *Proc. International Conference on Machine Learning (ICML)*. 44, 45

Kok, S. and Domingos, P. (2007). Statistical predicate invention. In *Proc. International Conference on Machine Learning (ICML)*, 433–440. Corvallis, OR. 48

Kok, S. and Domingos, P. (2010). Learning Markov logic networks using structural motifs. In *Proc. International Conference on Machine Learning (ICML)*. 45

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models - Principles and Techniques*. MIT Press. 7, 99, 161

Koller, D. and Pfeffer, A. (1997). Object-oriented Bayesian networks. In *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*, 302–313. 319

Kondor, R. and Lafferty, J. D. (2002). Diffusion kernels on graphs and other discrete input spaces. In *Proc. International Conference on Machine Learning (ICML)*, 315–322. Morgan Kaufmann. 368

Kopp, T., Singla, P., and Kautz, H. (2015). Lifted symmetry detection and breaking for map inference. In *Advances in Neural Information Processing Systems*, 1315–1323. 89

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37. 14

Kowalski, R. A. (2014). *Logic for Problem Solving, Revisited*. Books on Demand. 13

Kroc, L., Sabharwal, A., and Selman, B. (2007). Survey propagation revisited. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 217–226. 213

Kroc, L., Sabharwal, A., and Selman, B. (2009). Messagepassing and local heuristics as decimation strategies for satisfiability. In *Proceedings of the ACM Symposium on Applied Computing (SAC)*, 1408–1414. 213

Kroc, L., Sabharwal, A., and Selman, B. (2011). Leveraging belief propagation, backtrack search, and statistics for model counting. *Annals of Operations Research*, 184(1):209–231. 216

Kschischang, F. R., Frey, B. J., and Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519. 59, 66

Kumar, A. and Zilberstein, S. (2010). Map estimation for graphical models by likelihood maximization. In *Advances in Neural Information Processing Systems 23 (NIPS)*, 1180–1188. 212, 238, 239, 240, 242, 249, 252

Kumaraswamy, R., Odom, P., Kersting, K., Leake, D., and Natarajan, S. (2015). Transfer learning across relational and uncertain domains: A language-bias approach. In *ICDM*. 48, 51, 52

Kuzelka, O. and Kungurtsev, V. (2019). Lifted weight learning of markov logic networks revisited. *arXiv preprint arXiv:1903.03099*. 159

Kuzelka, O. and Wang, Y. (2020). Domain-liftability of relational marginal polytopes. *arXiv preprint arXiv:2001.05198*. 159

Lang, M. and Toussaint, M. (2009). Approximate inference for planning in stochastic relational worlds. In *Proc. International Conference on Machine Learning*. 376, 377

Laplace, P. S. (1814). *Essai philosophique sur les probabilities*. Courcier. Reprinted (1812) in English, F.W. Truscott amd F. L. Emory (Trans.) by Wiley, New York. 15

Larranaga, P., Kuijpers, C. M., Poza, M., and Murga, R. H. (1997). Decomposing Bayesian networks: triangulation of the moral graph with genetic algorithms. *Statistics and Computing*, 7(1):19–34. 99

Lasserre, J. (2002). An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM Journal on Optimization*, 12(3):756–769. 364

Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 157–224. 147, 165

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press. 161

Lauritzen, S. L., Barndorff-Nielsen, O. E., Dawid, A. P., Diaconis, P., and Johansen, S. (1984). Extreme point models in statistics. *Scandinavian Journal of Statistics*, 11(2). 165

Lee, J., Marinescau, R., and Dechter, R. (2014). Applying marginal map search to probabilistic conformant planning. In *Fourth International Workshop on Statistical Relational AI (StarAI)*. 376

Lee, J., Marinescau, R., and Dechter, R. (2016). Applying search based probabilistic inference algorithms to probabilistic conformant planning: Preliminary results. In *Proceedings of the International Symposium on Artificial Intelligence and Mathematics (ISAIM)*. 376

Leeuwen, J. (1990). *Handbook of theoretical computer science: Algorithms and complexity*, volume 1. Elsevier. 159

Lesner, B. and Zanuttini, B. (2011). Efficient policy construction for MDPs represented in probabilistic PDDL. In Bacchus, F., Domshlak, C., Edelkamp, S., and Helmert, M. (eds.), *ICAPS*. AAAI. 393

Levine, S. (2018). Reinforcement learning and control as probabilistic inference: Tutorial and review. *arXiv*, 1805.00909. 377

Li, W., Saidi, H., Sanchez, H., Schäf, M., and Schweitzer, P. (2016). Detecting similar programs via the weisfeiler-leman graph kernel. In *ICSR*, volume 9679 of *Lecture Notes in Computer Science*, 315–330. Springer. 370

Limketkai, B., Liao, L., and Fox, D. (2005). Relational object maps for mobile robots. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, IJCAI 2005*, 1471–1476. 318, 327

Liu, Q. and Ihler, A. T. (2012). Belief propagation for structured decision making. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 523–532. 376, 377

Lloyd, J. (1987). *Foundations of Logic Programming*. Springer Verlag. Second Edition. 381

Lovász, L. and Schrijver, L. (1991). Cones of matrices and set-functions and 0–1 optimization. *Cones of Matrices and Set-Functions and 0–1 Optimization SIAM Journal on Optimization*, 1(2):166–190. 364

Loveland, D. W. (1978). *Automated Theorem Proving: A Logical Basis*. Fundamental Studies in Computer Science. North-Holland. 13

Lowd, D. and Domingos, P. (2007). Efficient weight learning for Markov logic networks. In *Proceedings of the Eleventh European Conference on Principles and Practice of Knowledge Discovery in Databases*, 200–211. Springer, Warsaw, Poland. 44

Luby, M. and Vigoda, E. (1999). Fast convergence of the glauber dynamics for sampling independent sets. *Random Struct. Algorithms*, 15(3-4):229–241. 196

Madan, G., Anand, A., Singla, P., et al. (2018). Block-value symmetries in probabilistic graphical models. *arXiv preprint arXiv:1807.00643*. 198

Malkin, P. (2014). Sherali–adams relaxations of graph isomorphism polytopes. *Discrete Optimization*, 12:73–97. 364

Maneva, E., Mossel, E., and Wainwright, M. (2007). A new look at survey propagation and its generalizations. *Journal of the ACM (JACM)*, 54:2–41. 213

Marshall, A., Olkin, I., and Arnold, B. (2011). *Inequalities: Theory of majorization and its applications*. Springer series in statistics. Springer, New York. 302

McCarthy, J. (1958). Programs with common sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, volume 1, 77–84. National Physical Laboratory. Reprinted in R. Brachman and H. Levesque (Eds.), Readings in Knowledge Representation, 1985, Morgan Kaufmann, Los Altos, CA. 375

McKay, B. D. (1981). Practical graph isomorphism. *Congressus Numerantium*, 30:45–87. 353, 356

Meert, W., Van den Broeck, G., and Darwiche, A. (2014). Lifted inference for probabilistic logic programs. In *Workshop on Probabilistic Logic Programming (PLP)*. 101

Meert, W., Vlasselaer, J., and Van den Broeck, G. (2016). A relaxed tseitin transformation for weighted model counting. In *Proceedings of the Sixth International Workshop on Statistical Relational AI (StarAI)*, 1–7. 101, 143

Meshi, O., Jaimovich, A., Globerson, A., and Friedman, N. (2009). Convexifying the Bethe free energy. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, 402–410. 281, 303, 304

Mézard, M. and Montanari, A. (2009). *Information, Physics, and Computation*. Oxford University Press. 227, 229

Mézard, M., Parisi, G., and Zecchina, R. (2002). Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812–815. 213, 220

Mihalkova, L., Huynh, T., and Mooney, R. J. (2007). Mapping and revising Markov logic networks for transfer learning. In *Proc. AAAI National Conference on Artificial intelligence*, 608–614. Proc. AAAI Conference on Artificial Intelligence, Vancouver, Canada. 45, 48

Milch, B., Marthi, B., Russell, S. J., Sontag, D., Ong, D. L., and Kolobov, A. (2005). BLOG: Probabilistic models with unknown objects. In Kaelbling, L. P. and Saffiotti, A. (eds.), *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 1352–1359. Professional Book Center. 37, 40, 319, 327

Milch, B. and Russell, S. J. (2006). First-order probabilistic languages: Into the unknown. In *International Conference on Inductive Logic Programming (ILP)*. 318, 319, 320, 327

Milch, B., Zettlemoyer, L. S., Kersting, K., Haimes, M., and Kaelbling, L. P. (2008). Lifted probabilistic inference with counting formulas. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*, 1062–1608. 57, 58, 66, 68, 71, 73, 75, 76, 79, 81, 82, 85, 89, 101, 105, 107, 113, 118, 120, 125, 127, 174, 276, 319

Mladenov, M., Ahmadi, B., and Kersting, K. (2012). Lifted linear programming. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 788–797. 164, 208

Mladenov, M., Globerson, A., and Kersting, K. (2014). Lifted message passing as reparametrization of graphical models. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, 603–612. 256

Mladenov, M. and Kersting, K. (2013). Lifted inference via k-locality. In *Proceedings of the 3rd International Workshop on Statistical Relational AI*. 185

Mohan, K. and Pearl, J. (2014). Graphical models for recovering probabilistic and causal queries from missing data. In Welling, M., Ghahramani, Z., Cortes, C., and Lawrence, N. (eds.), *Advances of Neural Information Processing 27 (NIPS Proceedings)*, 1520–1528. 14

Montanari, A., Ricci-Tersenghi, F., and Semerjian, G. (2007). Solving constraint satisfaction problems through belief propagation-guided decimation. In *Proceedings of the 45th Allerton Conference on Communications, Control and Computing*, 352–359. 213, 216, 217, 227

Mooij, J. M. (2008). *Understanding and Improving Belief Propagation*. Ph.D. thesis, Radboud University Nijmegen. 252

Mooij, J. M. (2010). libDAI: A free and open source C++ library for discrete approximate inference in graphical models. *Journal of Machine Learning Research*, 11:2169–2173. 224, 253

Morettin, P., Passerini, A., and Sebastiani, R. (2019). Advanced SMT techniques for weighted model integration. *Artificial Intelligence*, 275:1–27. 101

Morik, K. and Kietz, J. (1989). A bootstrapping approach to concept clustering. In *Proceedings of the 6th International Workshop on Machine Learning (ML)*, 503–504. 239

Morris, C., Kriege, N. M., Kersting, K., and Mutzel, P. (2016). Faster kernels for graphs with continuous attributes via hashing. *CoRR*, abs/1610.00064. 370

Morris, C., Ritzert, M., Fey, M., Hamilton, W., Lenssen, J., rattan, G., and Grohe, M. (2019). Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 370

Muggleton, S. (1996). Stochastic logic programs. In *Advances in Inductive Logic Programming*, 254–264. 25, 35

Muggleton, S. and de Raedt, L. (1994). Special issue: Ten years of logic programming inductive logic programming: Theory and methods. *The Journal of Logic Programming*, 19:629 – 679. 40

Murphy, K., Weiss, Y., and Jordan, M. (1999). Loopy Belief Propagation for Approximate Inference: An Empirical Study. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence (UAI-99)*, 467–475. 206

Natarajan, S., Khot, T., Kersting, K., Gutmann, B., and Shavlik, J. (2012). Gradient-based boosting for statistical relational learning: The Relational Dependency Network case. *Machine Learning Journal*. 45, 46, 47

Natarajan, S., Khot, T., Kersting, K., and Shavlik, J. (2015). *Boosted Statistical Relational Learners: From Benchmarks to Data-Driven Medicine*. SpringerBriefs in Computer Science. 46

Natarajan, S., Tadepalli, P., Altendorf, E., Dietterich, T. G., Fern, A., and Restificar, A. C. (2005). Learning first-order probabilistic models with combining rules. In De Raedt, L. and Wrobel, S. (eds.), *Proceedings of the 22nd International Conference on Machine Learning (ICML-05)*, volume 119 of *ACM International Conference Proceeding Series*, 609–616. ACM. 10, 33, 41, 51

Natarajan, S., Tadepalli, P., Dietterich, T. G., and Fern, A. (2009). Learning first-order probabilistic models with combining rules. *Special Issue on Probabilistic Relational Learning, AMAI*. 36

Nath, A. and Domingos, P. (2010). Efficient lifting for online probabilistic inference. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 1193–1198. 228, 275, 341

Neville, J. and Jensen, D. (2007). Relational dependency networks. *Journal of Machine Learning Research*, 8:653–692. 12, 31, 40, 45

Ng, R. and Subrahmanian, V. S. (1992). Probabilistic logic programming. *Information and Computation*, 101(2):150–201. 319

Ngo, L. and Haddawy, P. (1995). Probabilistic logic programming and Bayesian networks. In *Proceedings ACSC95*. 39

Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33. 3

Niemira, M. P. and Saaty, T. L. (2004). An analytic network process model for financial-crisis forecasting. *International Journal of Forecasting*, 20(4):573–587. 318

Niepert, M. (2012a). Lifted probabilistic inference: An MCMC perspective. In *Proceedings of the 2nd International Workshop on Statistical Relational AI (StaRAI)*. 172, 198

Niepert, M. (2012b). Markov chains on orbits of permutation groups. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 624–633. 89, 164, 170, 174, 176, 184, 185, 186, 188, 189, 193, 195, 196, 198

Niepert, M. (2013). Symmetry-aware marginal density estimation. In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI)*. 164, 174, 184, 185

Niepert, M. and Domingos, P. (2014). Exchangeable variable models. In *Proceedings of the International Conference on Machine Learning (ICML)*. 174

Niepert, M. and Van den Broeck, G. (2014). Tractability through exchangeability: A new perspective on efficient probabilistic inference. In *AAAI*, 2467–2475. 175, 185

Nitti, D., De Laet, T., and De Raedt, L. (2013). A particle filter for hybrid relational domains. In Amato, N. (ed.), *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-13)*, 2764–2771. IEEE. 36

Noessner, J., Niepert, M., and Stuckenschmidt, H. (2013). RockIt: Exploiting Parallelism and Symmetry for MAP Inference in Statistical Relational Models. In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI)*. 164, 185

Paige, R. and Tarjan, R. (1987). Three partition refinement algorithms. *SIAM Journal on Computing*, 16(6):973–989. 350, 353

Pak, I. (2000). The product replacement algorithm is polynomial. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 476–485. 189

Pasula, H., Marthi, B., Milch, B., Russell, S. J., and Shpitser, I. (2003). Identity uncertainty and citation matching. In *Advances in Neural Information Processing Systems*, 1425–1432. 37, 38

Pearl, J. (1986). Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288. 117

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers. 3, 7, 10, 22, 26, 161, 174, 205, 263, 273

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition. 8, 14, 15

Pearson, M. and Michell, L. (2000). Smoke Rings: social network analysis of friendship groups, smoking and drug-taking. *Drugs: education, prevention and policy*, 7:21–37. 126

Pednault, E. P. D. (1989). ADL: Exploring the middle ground between STRIPS and the situation calculus. In *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 324–332. 374

Perlich, C. and Provost, F. J. (2003). Aggregation-based feature invention and relational concept classes. In *ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. 10, 33

Pfeffer, A. (2007). The design and implementation of IBAL: A general-purpose probabilistic language. In Getoor, L. and Taskar, B. (eds.), *Statistical Relational Learning*. MIT Press. 53

Pfeffer, A., Koller, D., Milch, B., and Takusagawa, K. T. (1999). Spook: A system for probabilistic object-oriented knowledge representation. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, UAI 1999*, 541–550. 39, 319

Piṅgala (200 BC). *Chandah-sûtra*. 123

Poole, D. (1991). Representing Bayesian networks within probabilistic Horn abduction. In *Proc. Seventh Conference on Uncertainty in Artificial Intelligence (UAI-91)*, 271–278. 15, 21

Poole, D. (1993). Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64:81–129. 3, 4, 15, 21, 25, 27, 35, 39

Poole, D. (1997). The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence*, 94:7–56. Special issue on economic principles of multi-agent systems. 21, 25

Poole, D. (2000). Abducing through negation as failure: Stable models within the independent choice logic. *Journal of Logic Programming*, 44(1–3):5–35. 21

Poole, D. (2003). First-order probabilistic inference. In Gottlob, G. and Walsh, T. (eds.), *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 985–991. Morgan Kaufmann Publishers. 17, 18, 57, 71, 79, 81, 89, 94, 101, 105, 113, 126, 267, 270, 275, 318, 319, 320, 327, 335

Poole, D. (2007). Logical generative models for probabilistic reasoning about existence, roles and identity. In *22nd AAAI Conference on AI (AAAI-07)*. 37

Poole, D. (2008). The independent choice logic and beyond. In *De Raedt et al (Eds) Probabilistic Inductive Logic Programming*, 222–243. Springer-Verlag. 37

Poole, D., Bacchus, F., and Kisynski, J. (2011). Towards completely lifted search-based probabilistic inference. *CoRR*, abs/1107.4035. 89, 105, 155, 156, 270

Poole, D., Buchman, D., Kazemi, S. M., Kersting, K., and Natarajan, S. (2014). Population size extrapolation in relational probabilistic modelling. In *Proc. of the Eighth International Conference on Scalable Uncertainty Management*, volume LNAI 8720, 292–305. 31, 32, 33

Poole, D. L. and Mackworth, A. K. (2017). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2nd edition. 3

Poon, H. and Domingos, P. (2006). Sound and efficient inference with probabilistic and deterministic dependencies. In Gil, Y. and Mooney, R. J. (eds.), *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, 458–463. AAAI Press. 190, 195, 212

Poon, H. and Domingos, P. (2011). Sum-product networks: A new deep architecture. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 689–690. IEEE. 149

Poon, H., Domingos, P., and Summer, M. (2008). A general method for reducing the complexity of relational inference and its application to MCMC. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*, 1075–1080. 212

Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York. 384

Raedt, L. D., Kimmig, A., and Toivonen, H. (2007). Problog: A probabilistic prolog and its application in link discovery. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, 2462–2467. 382

Raghavan, A., Joshi, S., Fern, A., Tadepalli, P., and Khardon, R. (2012). Planning in factored action spaces with symbolic dynamic programming. In *Proc. AAAI Conference on Artificial Intelligence*. 393

Raghavan, A., Khardon, R., Fern, A., and Tadepalli, P. (2013). Symbolic opportunistic policy iteration for factored-action MDPs. In *Proc. Advances in Neural Information Processing Systems*, 2499–2507. 393

Ramakrishnan, R. and Gehrke, J. (2003). *Database management systems (3. ed.)*. McGraw-Hill. 66

Ramanan, N., Kunapuli, G., Khot, T., Fatemi, B., Kazemi, S. M., Poole, D., Kersting, K., and Natarajan, S. (2018). Structure learning for relational logistic regression: An ensemble approach. In *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*. 40

Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., and Rao, C. R. (1973). *Linear statistical inference and its applications*, volume 2. Wiley New York. 13

Ravkic, I., Ramon, J., and Davis, J. (2015). Learning relational dependency networks in hybrid domains. *Machine Learning*, 100(2-3):217–254. 47

Read, R. and Corneil, D. (1977). The graph isomorphism disease. *Journal of Graph Theory*, 1(4):339–363. 350

Reiter, R. (2001). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press. 385

Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Duxbury Press. 129

Richards, B. and Mooney, R. (1995). Automated refinement of first-order horn-clause domain theories. *Machine Learning*, 19(2):95–131. 51

Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine Learning*, 62(1–2):107–136. 17, 19, 20, 40, 41, 42, 44, 51, 52, 101, 137, 139, 163, 206, 250, 315, 318, 319, 327, 382

Riguzzi, F., Bellodi, E., Zese, R., Cota, G., and Lamma, E. (2017). A survey of lifted inference approaches for probabilistic logic programming under the distribution semantics. *International Journal of Approximate Reasoning*, 80:313–333. 101

Ristoski, P. and Paulheim, H. (2016). RDF2Vec: RDF graph embeddings for data mining. In *International Semantic Web Conference (1)*, volume 9981 of *Lecture Notes in Computer Science*, 498–514. 370

Robertson, N. and Seymour, P. (1986). Graph minors. II. Algorithmic aspects of tree-width. *Journal of algorithms*, 7(3):309–322. 161

Rose, D. J., Tarjan, R. E., and Lueker, G. S. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on computing*, 5(2):266–283. 99

Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3rd edition. 3, 381, 384

Sahs, J. and Khan, L. (2012). A machine learning approach to android malware detection. In *EISIC*, 141–147. IEEE Computer Society. 370

Sanner, S. (2008). *First-order Decision-theoretic Planning in Structured Relational Environments*. Ph.D. thesis, University of Toronto, Toronto, ON, Canada. 386, 390, 393

Sanner, S. and Boutilier, C. (2005). Approximate linear programming for first-order MDPs. In *Uncertainty in Artificial Intelligence (UAI-05)*, 509–517. Edinburgh, Scotland. 394

Sanner, S. and Boutilier, C. (2006). Practical linear evaluation techniques for first-order MDPs. In *Uncertainty in Artificial Intelligence (UAI-06)*. Boston, Mass. 394

Sanner, S. and Boutilier, C. (2007). Approximate solution techniques for factored first-order MDPs. In *Proceedings of the Seventeenth International Conference on Automated Planning and Scheduling, ICAPS 2007*, 288–295. 386, 390

Sanner, S. and Boutilier, C. (2009). Practical solution techniques for first-order MDPs. *Artif. Intell.*, 173:748–488. 381, 384, 393, 394

Sanner, S. and Kersting, K. (2010). Symbolic dynamic programming for first-order POMDPs. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. 394

Sarkhel, S., Venugopal, D., Singla, P., and Gogate, V. G. (2014). An integer polynomial programming based framework for lifted map inference. In *Advances in Neural Information Processing Systems*, 3302–3310. 106

Sato, N. and Tinney, W. F. (1963). Techniques for exploiting the sparsity of the network admittance matrix. *Power Apparatus and Systems, IEEE Transactions on*, 82(69):944–950. 99

Sato, T. (1995). A statistical learning method for logic programs with distribution semantics. In *Proceedings of the 12th International Conference on Logic Programming (ICLP95)*, 715–729. 15, 21, 25, 141

Sato, T. and Kameya, Y. (1997). PRISM: A symbolic-statistical modeling language. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, 1330–1335. 3

Sato, T. and Kameya, Y. (2001). Parameter learning of logic programs for symbolic-statistical modeling. *Journal of Artificial Intelligence Research (JAIR)*, 15:391–454. 21, 25, 35, 36, 40

Sato, T., Kameya, Y., and Zhou, N.-F. (2005). Generative modeling with failure in PRISM. In Kaelbling, L. P. and Saffiotti, A. (eds.), *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 847–852. Professional Book Center. 21

Saul, L. K., Jaakkola, T., and Jordan, M. I. (1996). Mean field theory for sigmoid belief networks. *arXiv preprint cs/9603102*. 28

Savicky, P. and Vomlel, J. (2007). Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika*, 43(5):747–764. 121

Selman, B., Kautz, H., and Cohen, B. (1995). Local search strategies for satisfiability testing. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. 235

Sen, P., Deshpande, A., and Getoor, L. (2009a). Bisimulation-based approximate lifted inference. In *Proceedings of the twenty-fifth Conference on Uncertainty in Artificial Intelligence*, 496–505. AUAI Press. 57, 274

Sen, P., Deshpande, A., and Getoor, L. (2009b). PrDB: managing and exploiting rich correlations in probabilistic databases. *VLDB Journal*, 18(5):1065–1090. 185

Shafer, G. R. and Shenoy, P. P. (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence*, 2(1-4):327–351. 147

Shariff, R., György, A., and Szepesvári, C. (2015). Exploiting symmetries to construct efficient mcmc algorithms with an application to slam. In *Artificial Intelligence and Statistics*, 866–874. 198

Sheldon, D. and Dietterich, T. (2011). Collective graphical models. In *Advances in Neural Information Processing Systems (NIPS)*, 1161–1169. 174

Shental, O., Bickson, D., Siegel, P. H., Wolf, J. K., and Dolev, D. (2008). Gaussian belief propagation solver for systems of linear equations. In *IEEE Int. Symp. on Inform. Theory (ISIT)*. Toronto, Canada. 337, 338

Sherali, H. D. and Adams, W. P. (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430. 364

Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M. (2011). Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561. 369, 370

Singla, P. and Domingos, P. (2006a). Entity resolution with Markov logic. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM-06)*, 572–582. 37

Singla, P. and Domingos, P. (2006b). Memory-efficient inference in relational domains. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 488–493. 212

Singla, P. and Domingos, P. (2008). Lifted first-order belief propagation. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*, 1094–1099. 89, 128, 185, 195, 206, 210, 211, 220, 260, 274, 276, 277, 304, 319, 327

Singla, P., Nath, A., and Domingos, P. (2010). Approximate Lifted Belief Propagation. In *Proceedings of the 1st International Workshop on Statistical Relation AI (StaRAI)*, 92–97. 248, 275

Singla, P., Nath, A., and Domingos, P. (2014). Approximate lifting techniques for belief propagation. In *Proceedings of AAAI*. 194

Smith, D. B. and Gogate, V. G. (2015). Bounding the cost of search-based lifted inference. In *Advances in Neural Information Processing Systems*, 946–954. 110

Sontag, D., Globerson, A., and Jaakkola, T. (2008a). Clusters and coarse partitions in LP relaxations. In *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS)*, 1537–1544. 305

Sontag, D., Meltzer, T., Globerson, A., Jaakkola, T., and Weiss, Y. (2008b). Tightening LP relaxations for map using message passing. In *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence (UAI)*, 503–510. 305

St-Aubin, R., Hoey, J., and Boutilier, C. (2000). APRICODD: Approximate policy construction using decision diagrams. In *Advances in Neural Information Processing 13 (NIPS-00)*, 1089–1095. Denver. 393

Suciu, D., Olteanu, D., Ré, C., and Koch, C. (2011). *Probabilistic Databases*, volume 16 of *Synthesis Lectures on Data Management*. Morgan & Claypool Publishers. 101, 106, 143

Taghipour, N. and Davis, J. (2012). Generalized counting for lifted variable elimination. In *Proceedings of the 2nd International Workshop on Statistical Relational AI (StaRAI)*, 1–8. 57, 174

Taghipour, N., Davis, J., and Blockeel, H. (2013a). Generalized counting for lifted variable elimination. In *International Conference on Inductive Logic Programming*, 107–122. Springer. 57

Taghipour, N., Fierens, D., Davis, J., and Blockeel, H. (2013b). Lifted variable elimination: Decoupling the operators from the constraint language. *Journal of Artificial Intelligence Research*, 47:393–439. 57

Taghipour, N., Fierens, D., Van den Broeck, G., Davis, J., and Blockeel, H. (2013c). Completeness results for lifted variable elimination. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*. (Under review). 57, 89, 105, 109, 164

Tarjan, R. E. and Mihalis, Y. (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on computing*, 13(3):566–579. 99

Taskar, B., Abbeel, P., and Koller, D. (2002). Discriminative Probabilistic Models for Relational Data. In Darwiche, A. and Friedman, N. (eds.), *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*, 485–492. 40

Thon, I., Landwehr, N., and De Raedt, L. (2011). Stochastic relational processes: Efficient inference and applications. *Machine Learning*, 82(2):239–272. 36

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728. 190

Tinhofer, G. (1991). A note on compact graphs. *Discrete Applied Mathematics*, 30:253–264. 362

Toussaint, M., Charlin, L., and Poupart, P. (2008). Hierarchical POMDP controller optimization by likelihood maximization. In *Proceedings of the 24th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 562–570. 238, 241

Toussaint, M. and Storsky, A. (2006). Probabilistic inference for solving discrete and continuous sta te Markov decision processes. In *Proc. International Conference on Machine Learning*. 376, 377

Tseitin, G. (1968). On the complexity of derivation in propositional calculus. *Studies in Constrained Mathematics and Mathematical Logic*. 96, 101

Valiant, L. G. (1979). The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421. 159

van Bremen, T. and Kuzelka, O. (2020). Approximate weighted first-order model counting: Exploiting fast approximate model counters and symmetry. *arXiv preprint arXiv:2001.05263*. 101

van de Meent, J., Paige, B., Tolpin, D., and Wood, F. (2016). Black-box policy search with probabilistic programs. In *Proceedings of the International Conference on Artificial Intelligence and Statistics, AISTATS*, 1195–1204. 376, 377

Van den Broeck, G. (2011). On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Proceedings of the 24th Annual Conference on Advances in Neural Information Processing Systems(NIPS)*, 1386–1394. 109, 155, 156, 164, 264, 276, 392

Van den Broeck, G. (2013). *Lifted Inference and Learning in Statistical Relational Models*. Ph.D. thesis, KU Leuven. 100, 103, 113, 147, 272, 392

Van den Broeck, G. (2015). Towards high-level probabilistic reasoning with lifted inference. In *Proceedings of the AAAI Spring Symposium on KRR*. xvi, 100

Van den Broeck, G. (2016). First-order model counting in a nutshell. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), Early Career Spotlight Track*. 11

Van den Broeck, G., Choi, A., and Darwiche, A. (2012). Lifted relax, compensate and then recover: From approximate to exact lifted probabilistic inference. In *Conference on Uncertainty in Artificial Intelligence (UAI)*. 239

Van den Broeck, G. and Darwiche, A. (2013). On the complexity and approximation of binary evidence in lifted inference. In *Advances in Neural Information Processing Systems*, 2868–2876. 89, 172, 192, 194, 197, 254

Van den Broeck, G. and Davis, J. (2012). Conditioning in first-order knowledge compilation and lifted probabilistic inference. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, 1–7. 149, 173

Van den Broeck, G., Meert, W., and Darwiche, A. (2014). Skolemization for weighted first-order model counting. In Baral, C., Giacomo, G. D., and Eiter, T. (eds.), *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR-14)*, 111–120. AAAI Press. 89, 103, 104, 139, 155, 156, 164

Van den Broeck, G., Meert, W., and Davis, J. (2013). Lifted generative parameter learning. In *Statistical Relational AI (StaRAI) workshop*. 159

Van den Broeck, G. and Niepert, M. (2015). Lifted probabilistic inference for asymmetric graphical models. In *AAAI*, 3599–3605. 198

Van den Broeck, G. and Suciu, D. (2017). Query processing on probabilistic data: A survey. *Foundations and Trends® in Databases*, 7(3-4):197–341. 95, 101

Van den Broeck, G., Taghipour, N., Meert, W., Davis, J., and De Raedt, L. (2011). Lifted probabilistic inference by first-order knowledge compilation. In Walsh, T. (ed.), *Proc. International Joint Conference on AI (IJCAI)*, 2178–2185. AAAI Press. 89, 90, 101, 105, 140, 147, 149, 174, 275, 276, 382

Van Gelder, A., Ross, K. A., and Schlipf, J. S. (1991). The well-founded semantics for general logic programs. *Journal of the ACM (JACM)*, 38(3):619–649. 141

Van Haaren, J., Kolobov, A., and Davis, J. (2015). Todtler: Two-order-deep transfer learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 3007–3015. 48, 49

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. 155

Vennekens, J., Denecker, M., and Bruynooghe, M. (2009). CP-logic: A language of causal probabilistic events and its relation to logic programming. *Theory and Practice of Logic Programming (TPLP)*, to appear. 25, 26

Vennekens, J., Verbaeten, S., and Bruynooghe, M. (2004). Logic programs with annotated disjunctions. In Demoen, B. and Lifschitz, V. (eds.), *Proceedings of the 20th International Conference on Logic Programming (ICLP-04)*, volume 3132 of *Lecture Notes in Computer Science*, 431–445. Springer. 25, 26

Venugopal, D. and Gogate, V. (2012). On lifting the gibbs sampling algorithm. In *Proceedings of the 26th Annual Conference on Advances in Neural Information Processing Systems (NIPS)*, 1–6. 185

Venugopal, D. and Gogate, V. (2014a). Evidence-based clustering for scalable inference in Markov logic. In *ECML PKDD*, 258–273. 89, 194

Venugopal, D. and Gogate, V. G. (2014b). Scaling-up importance sampling for Markov logic networks. In *Advances in Neural Information Processing Systems*, 2978–2986. 89

Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., and Borgwardt, K. M. (2010). Graph kernels. *Journal of Machine Learning Research*, 11:1201–1242. 368

Vlasselaer, J., Kimmig, A., Dries, A., Meert, W., and De Raedt, L. (2016a). Knowledge compilation and weighted model counting for inference in probabilistic logic programs. In *AAAI Workshop: Beyond NP*. 101

Vlasselaer, J., Meert, W., Van den Broeck, G., and De Raedt, L. (2016b). Exploiting local and repeated structure in dynamic bayesian networks. *Artificial Intelligence*, 232:43 – 53. 95

Wainwright, M., Jaakkola, T., and Willsky, A. (2005). Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717. 305, 306

Wainwright, M. and Jordan, M. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1-2):1–305. 281, 285, 286

Wang, C., Joshi, S., and Khardon, R. (2008). First order decision diagrams for relational MDPs. *Journal of Artificial Intelligence Research (JAIR)*, 31:431–472. 385, 387, 393

Wang, C. and Khardon, R. (2007). Policy iteration for relational MDPs. In *Uncertainty in Artificial Intelligence (UAI-07)*. Vancouver, Canada. 394

Wang, C. and Khardon, R. (2010). Relational partially observable MDPs. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. 394

Wang, J. and Domingos, P. (2008). Hybrid Markov logic networks. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008*. 327

Weiss, Y. and Freeman, W. (2001). Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, 13(10):2173–330. 337

Wellman, M. P., Breese, J. S., and Goldman, R. P. (1992). From knowledge bases to decision models. *The Knowledge Engineering Review*, 7(1):35–53. 19

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2003). Understanding belief propagation and its generalizations. In Lakemeyer, G. and Nebel, B. (eds.), *Exploring Artificial Intelligence in the New Millennium*, chapter 8, 239–269. Morgan Kaufmann. 263, 273

Yoon, S., Fern, A., and Givan, R. (2002). Inductive policy selection for first-order Markov decision processes. In *Uncertainty in Artificial Intelligence (UAI-02)*, 569–576. Edmonton. 394

Yoon, S., Fern, A., and Givan, R. (2006). Approximate policy iteration with a policy language bias: Learning to solve relational Markov decision processes. *Journal of Artificial Intelligence Research (JAIR)*, 25:85–118. 394

Younes, H. L. S., Littman, M. L., Weissman, D., and Asmuth, J. (2005). The first probabilistic track of the international planning competition. *Journal of Artificial Intelligence Research (JAIR)*, 24:851–887. 393

Zeng, Z. and Van den Broeck, G. (2019). Efficient search-based weighted model integration. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*. 101

Zhang, H. and Stickel, M. E. (1996). An efficient algorithm for unit propagation. In *Proceedings of the 4th International Symposium on Artificial Intelligence and Mathematics (AI-MATH)*, 166–169. 217

Zhang, N. L. and Poole, D. (1994). A simple approach to Bayesian network computations. In *Proceedings of the 10th Canadian Conference on AI*, 171–178. 12, 115

Zhang, N. L. and Poole, D. (1996). Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research (JAIR)*, 5:301–328. 117