On Computing Total Variation Distance between Markov Logic Networks

Zixiang Xiong XIONGZIXIANG3619@163.COM
CRRC Zhuzhou Institute
zhuzhou, China

Peter Jung JUNGPETE@FEL.CVUT.CZ
Faculty of Electrical Engineering, Czech Technical University in Prague
Prague, Czech Republic

Qiaolan Meng MENGQL1606@BUAA.EDU.CN
Beihang University
Beijing, China

Yuanhong Wang LUCIENWANG@JLU.EDU.CN
Jilin University
Changchun, China

Ondřej Kuželka ONDREJ.KUZELKA@FEL.CVUT.CZ
Faculty of Electrical Engineering, Czech Technical University in Prague
Prague, Czech Republic

Yuyi Wang YUYIWANG920@GMAIL.COM
CRRC Zhuzhou Institute & Tengen Intelligence Institute
zhuzhou, China

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

The total variation (TV) distance is a fundamental metric for quantifying the difference between probability distributions. Bhattacharyya et al. (2024) showed that computing the TV distance between two directed graphical models is, in general, intractable. We extend this line of research to Markov Logic Networks (MLNs), a prominent class of statistical relational models that combine first-order logic with probabilistic graphical models. Specifically, we establish both tractability and hardness results regarding the computation of the TV distance between two MLNs. On the positive side, we reduce TV distance computation between MLNs to probabilistic inference, enabling the use of lifted inference techniques. Thereby, we identify classes of MLNs for which TV distance computation is tractable. On the negative side, we prove that, in general, the TV distance computation between two MLNs is intractable. In addition to these theoretical results, we empirically validate that soft rules with moderate weights can approximate hard rules, and we also present an application of the MLNs TV distance in benchmarking graph neural networks on synthetic data.

Keywords: TV Distance, Markov Logic Networks, Lifted Inference, Weighted First-Order Model Counting

© 2026 Z. Xiong, P. Jung, Q. Meng, Y. Wang, O. Kuželka & Y. Wang.

1. Introduction

The total variation (TV) distance measures the maximum discrepancy between the probabilities assigned by two distributions to the same event. It is one of the most widely used distance metrics in machine learning (Shalev-Shwartz and Ben-David, 2014), information theory (Cover and Thomas, 2006), differential privacy (Dwork, 2006) and randomized algorithms (Mitzenmacher and Upfal, 2005). The exact computation of the TV distance is known to be intractable, as it is #P-complete even for simple product distributions (Bhattacharyya et al., 2023). This result directly implies the hardness of computing the TV distance between two directed graphical models (Bhattacharyya et al., 2025). Notably, however, there remain important cases where the computation of the TV distance is tractable, ranging from binomial distributions to more intriguing random graph models, such as the Erdös-Rényi (ER) model. This observation motivates the question: *Is there a broader class of distributions for which the TV distance can be computed efficiently, and if so, what are the precise tractability boundaries?*

In this paper, we address this question by focusing on Markov Logic Networks (MLNs) (Richardson and Domingos, 2006), a prominent class of distributions originating from statistical relational models (SRMs) (Milch et al., 2008; Van den Broeck et al., 2021; Gehrke et al., 2019). An MLN is a set of weighted first-order logic formulas. Each formula represents a soft constraint on the relational structure, and its weight indicates the strength of the constraint. Given a domain, an MLN defines a distribution over its structures that can be viewed as a Markov random field. For instance, ER graphs can be represented as an MLN $\{(\forall x \forall y\colon (E(x,y) \rightarrow E(y,x)) \wedge \neg E(x,x), \infty), (E(x,y), \varpi)\}$ over a domain of vertices, where the first formula with infinite weight enforces the symmetry and irreflexivity of the edge relation, and the second formula with finite weight ϖ controls the density of the graph.

We use the theoretical framework of *domain-liftability* (Van den Broeck et al., 2011; Beame et al., 2015) to establish precise tractability boundaries for computing the TV distance between two MLNs. Domain-liftability, first introduced for probabilistic inference in MLNs, describes the complexity of inference as a function of the domain size, with the formulas and their weights being fixed. From the perspective of probabilistic graphical models, the weighted formulas in an MLN can be viewed as templates for generating a Markov random field over a domain, and domain-liftability characterizes the complexity of inference as the domain size grows while the templates remain unchanged (Richardson and Domingos, 2006). It is very natural to study domain-liftability in practice, as the formulas and their weights are often learned from data and thus fixed, while the domain size can vary significantly across different applications (Van Haaren et al., 2016; Malhotra et al., 2025).

There are several lifted inference approaches for MLNs, such as lifted variable elimination (Milch et al., 2008), lifted belief propagation (Taghipour et al., 2013), lifted junction tree algorithm (Braun and Möller, 2016; Gehrke et al., 2018), and weighted first-order model counting approaches (Van den Broeck et al., 2011; Beame et al., 2015; Kuzelka, 2021). The basic idea of these approaches is to perform inference by grouping together indistinguishable individuals in the domain, thus avoiding the combinatorial explosion of grounding the MLN into a propositional Markov random field. Our positive results utilize a similar idea and rely on a reduction from TV distance computation to WFOMC. The reduction would also apply to other lifted inference approaches, and we choose WFOMC for its generality

and the extensive literature on its complexity. The existing results on the complexity of WFOMC allow us to identify a wide range of MLNs for which the TV distance computation is tractable, e.g., those with formulas containing at most two logical variables (Van den Broeck et al., 2011).

Our negative results also build on the literature on the complexity of WFOMC, where it has been shown that WFOMC is intractable for sentences with formulas containing three or more logical variables (Beame et al., 2015). The reduction, now from WFOMC to TV distance computation, is proved for any fragment of first-order logic that allows arbitrary boolean combinations of formulas, which includes the three-variable fragment as a special case. As a result, we show that TV distance computation is intractable for MLNs involving formulas with three or more logical variables, establishing a precise boundary for the tractability of TV distance computation between MLNs.

Finally, with the positive result, we validate the extent to which a folk assumption - that hard rules in MLNs can be effectively approximated by soft rules with moderately large weights - actually holds in practice. Then we provide a new benchmarking approach for the graph classification task. In brief, we define two classes of graphs using MLNs from a tractable fragment, and use their TV distance to compute the optimal Bayes error rate for classifying graphs from these two classes. We demonstrate the effectiveness of this approach through experiments with graph neural networks (GNNs). This is particularly relevant as, in theory, GNNs are expressive enough to model the constraints due to their connection to first-order logics (Barceló et al., 2020).

2. Related Work

The computational complexity of total variation (TV) distance has recently attracted attention in the context of probabilistic graphical models. Bhattacharyya et al. (2023) studied the complexity of computing and approximating TV distance for product distributions and related probabilistic representations. They demonstrated that exactly computing the TV distance between product distributions is #P-hard and proved that deciding whether the TV distance between two Bayes nets of degree 2 is equal to 0 is NP-hard. Additionally, they provided a fully polynomial-time approximation scheme (FPTAS) for approximating the TV distance between an arbitrary product distribution and a product distribution with a constant number of distinct marginals, including the uniform distribution. Subsequently, Feng et al. (2023) presented a fully polynomial randomized approximation scheme (FPRAS) for approximating the TV distance between two arbitrary product distributions, and Feng et al. (2024) later gave an FPTAS for the same task. Both types of schemes run in time polynomial in the input size and $1/\varepsilon$; an FPRAS provides this guarantee with a constant success probability, whereas an FPTAS is deterministic. More recent work established a direct connection between TV distance estimation and probabilistic inference, giving a reduction from relative approximation of TV distance to inference in directed graphical models (Bhattacharyya et al., 2024). In this paper, we aim to extend this line of research to Markov Logic Networks (MLNs) and investigate the tractability of the exact computation of TV distance over MLNs.

Markov Logic Networks combine first-order logic with Markov networks by assigning weights to first-order formulas (Richardson and Domingos, 2006). They are a central formal-

ism in statistical relational learning, where probabilistic models are defined over individuals and relations rather than over a fixed set of propositional variables (Getoor and Taskar, 2007; Van den Broeck et al., 2021). Exact inference in grounded MLNs is generally intractable, motivating lifted inference methods that exploit symmetries and exchangeability among domain elements (Poole, 2003; De Salvo Braz et al., 2005; Kersting, 2012). Representative lifted inference approaches include lifted variable elimination, counting formulas, lifted belief propagation, lifted junction-tree methods, and first-order knowledge compilation (Milch et al., 2008; Singla and Domingos, 2008; Braun and Möller, 2016; Van den Broeck et al., 2011). *Weighted first-order model counting (WFOMC)* is particularly relevant to our work because inference in many first-order probabilistic models can be reduced to weighted model counting over first-order theories (Van den Broeck, 2016; Van den Broeck et al., 2014). A rich line of work has characterized the domain-liftability of WFOMC and lifted inference, including tractable fragments such as two-variable first-order logic and hardness results for more expressive fragments (Van den Broeck, 2011; Nima Taghipour et al., 2013; Beame et al., 2015; Kuzelka, 2021; Kuželka, 2023). These results provide the technical foundation for our tractability and hardness results.

3. Background

We use the standard definition of the total variation distance between two probability distributions. Given two probability distributions P and Q over the same measurement space S, the *total variation distance* (Bhattacharyya et al., 2023) (TV distance) between P and Q is defined as

$$d_{TV}(P,Q) := \max_{A \subseteq S} |P(A) - Q(A)| = \frac{1}{2} \sum_{x \in S} |P(x) - Q(x)|.$$

In the rest of this section, we review the other key concepts underlying our study, including the definition of MLNs, the weighted first-order model counting and distributions, and the domain-liftability framework for characterizing the complexity of inference in MLNs.

3.1. Markov Logic Networks

We consider *pure relational first-order logic without function symbols* defined by a set of constants Δ, a set of variables $\mathcal{V}$ and for each $k \in \mathbb{N}$ a set $\mathcal{P}_k$ of k-ary predicates. An atom is $P(x_1, \ldots, x_k)$ with $P \in \mathcal{P}_k$ and $x_1, \ldots, x_k \in \Delta \cup \mathcal{V}$. A literal is either an atom or its negation. A formula is recursively defined as an atom, the negation of another formula, or the conjunction or disjunction of two formulas. Additionally, a formula may be enclosed by one or more quantifiers, such as $\forall x$ or $\exists x$, and still be considered a formula. A logical variable in a formula is considered free if it is not bound by any quantifier. A formula devoid of free variables is called a sentence. Given a predicate vocabulary $\mathcal{P}$ and a domain Δ, an interpretation I maps each k-ary predicate $P/k \in \mathcal{P}$ to Δ^k, i.e. $I(P/k) \subseteq \Delta^k$. We call the set of all interpretations an interpretation space and denote it by $\mathcal{I}_{\mathcal{P},\Delta}$. Denote by $\mathcal{P}_\alpha$ the predicate vocabulary of a formula α. The satisfaction relation of a formula α and an interpretation $I \in \mathcal{I}_{\mathcal{P}_\alpha,\Delta}$ is defined in the usual way: $I \models \alpha$ means that the formula α is true in the interpretation I. An interpretation I is a model of a formula α if $I \models \alpha$. The set of all models of a formula α over a domain Δ is denoted by $\mathcal{M}_{\alpha,\Delta}$.

An MLN is a set of weighted first-order logic formulas (α, ϖ), where α is a first-order logic formula and $\varpi \in \mathbb{R} \cup \{\infty\}$ is weight. Let $\Phi^{\mathbb{R}}$ and Φ^{∞} be the subsets of (α, ϖ) in Φ with finite and infinite weights, respectively. Denote by $\phi_{\Phi}^{\infty} = \bigwedge_{(\alpha_i, \varpi_i) \in \Phi^{\infty}} \forall \mathbf{x}_i : \alpha_i(\mathbf{x}_i)$ the conjunction of all hard constraints in Φ, where $\mathbf{x}_i$ is the set of free variables in α_i.

Definition 1 *Given a domain Δ, an MLN Φ defines a distribution $D_{\Phi,\Delta}$ over $\mathcal{I}_{\mathcal{P}_\Phi,\Delta}$:*

$$D_{\Phi,\Delta}(I) := \begin{cases} \frac{1}{Z} \exp\left(\sum_{(\alpha,\varpi)\in\Phi^{\mathbb{R}}} \varpi \cdot n(\alpha, I)\right) & \text{if } I \models \phi_{\Phi}^{\infty}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{P}_\Phi$ is the predicate vocabulary of Φ, $n(\alpha, I)$ is the number of groundings of α satisfied in I (when α is a sentence, we define $n(\alpha, I) = \mathbb{1}[I \models \alpha]$), and Z is the normalization constant, i.e., the partition function of Φ over Δ.

3.2. Weighted First-Order Model Counting and Domain-Liftability

We first need the notion of first-order model counting (FOMC).

Definition 2 (First-Order Model Counting) *Given a sentence ϕ and a domain Δ, the* first-order model counting problem (FOMC) *is to compute* $\mathsf{FOMC}(\phi, \Delta) = |\mathcal{M}_{\phi,\Delta}|$*, where $\mathcal{M}_{\phi,\Delta}$ is the set of all models of ϕ over Δ.*

Definition 3 *Given a predicate vocabulary $\mathcal{P}$ and a pair of weighting functions $(w, \bar{w}) : \mathcal{P} \to \mathbb{R}_{\geq 0}$, the weight of an interpretation I is defined as*

$$W_{w,\bar{w}}(I) := \prod_{l \in I_T} w(\mathsf{pred}(l)) \cdot \prod_{l \in I_F} \bar{w}(\mathsf{pred}(l)), \tag{1}$$

where I_T and I_F are the sets of positive and negative literals interpreted as true in I, respectively, and $\mathsf{pred}(l)$ is the predicate symbol of the literal l.

Weighted first-order model counting (WFOMC) is a generalization of FOMC, which assigns a weight to each interpretation and computes the sum of weights of all models of a sentence.

Definition 4 (Weighted First-Order Model Counting) *Given a sentence ϕ, a domain Δ, and a pair of weighting functions $(w, \bar{w})$, the* weighted first-order model counting problem (WFOMC) *is to compute* $\mathsf{WFOMC}(\phi, \Delta, w, \bar{w}) := \sum_{I \in \mathcal{M}_{\phi,\Delta}} W_{w,\bar{w}}(I)$.

Let us illustrate the connection between WFOMC and probabilistic inference in MLNs via the reduction from the partition function computation of an MLN Φ to WFOMC. For each formula $(\alpha, \varpi) \in \Phi^{\mathbb{R}}$, we introduce a new predicate ξ and construct a sentence $\phi_{\Phi}^{\mathbb{R}}$ by taking the conjunction of all sentences $\forall \mathbf{x} : \xi(\mathbf{x}) \Leftrightarrow \alpha(\mathbf{x})$, where $\mathbf{x}$ are the free variables in α. Let $(w, \bar{w})$ be the weighting functions defined as $w(\xi) = e^{\varpi}$ and $\bar{w}(\xi) = 1$ for each ξ corresponding to a formula in $\Phi^{\mathbb{R}}$, and $w(P) = \bar{w}(P) = 1$ for all other predicates P. Then the partition function of Φ over Δ can be computed as $\mathsf{WFOMC}(\phi_{\Phi}^{\mathbb{R}} \wedge \phi_{\Phi}^{\infty}, \Delta, w, \bar{w})$.

Domain-liftability is defined in terms of the complexity of WFOMC, where a sentence ϕ is said to be domain-liftable if $\mathsf{WFOMC}(\phi, \Delta, w, \bar{w})$ can be computed in time polynomial

in the domain size $|\Delta|$ for any *fixed* pair of weighting functions $(w, \bar{w})$. A class of sentences (also called a fragment of first-order logic) is domain-liftable if every sentence in the class is domain-liftable. There are several known domain-liftable fragments, such as the two-variable fragment of first-order logic ($\mathbf{FO}^2$, restricted to at most two distinct variables) (Van den Broeck et al., 2011) and the two-variable fragment with counting quantifiers ($\mathbf{C}^2$, extending $\mathbf{FO}^2$ with $\exists^{\geq m}$) (Kuzelka, 2021). There are also intractable fragments, e.g., the three-variable fragment of first-order logic ($\mathbf{FO}^3$, allowing up to three variables) (Beame et al., 2015), where there is a sentence such that its WFOMC is $\#\mathsf{P}_1$-hard.[1]

3.3. Weighted First-Order Model Distributions

To align with WFOMC and the domain-liftability complexity, we use the *weighted first-order model distribution (WFOMD)* to represent the distributions defined by MLNs.

Definition 5 (Weighted First-Order Model Distribution) *Given a* satisfiable *first-order sentence ϕ, a domain Δ, and a pair of weighting functions $(w, \bar{w}) : P \to \mathbb{R}_{\geq 0}$, the* weighted first-order model distribution (WFOMD) *of ϕ over Δ under $(w, \bar{w})$, denoted as $D_{\phi,\Delta,w,\bar{w}}$, is defined over $\mathcal{I}_{\mathcal{P}_\phi,\Delta}$ as*

$$D_{\phi,\Delta,w,\bar{w}}(I) := \begin{cases} W_{w,\bar{w}}(I)/Z & \text{if } I \models \phi, \\ 0 & \text{otherwise}, \end{cases}$$

for all $I \in \mathcal{I}_{\mathcal{P}_\phi,\Delta}$, where $Z = \sum_{I' \in \mathcal{M}_{\phi,\Delta}} W_{w,\bar{w}}(I')$ is the normalization constant also known as the partition function.

Following the same idea as the reduction from partition function computation of an MLN to WFOMC, we can also reduce the distribution defined by an MLN over a domain to a WFOMD over the same domain under appropriate weighting functions.

Let ϕ_1 and ϕ_2 be two sentences that share the same predicate vocabulary. We say that (ϕ_1, ϕ_2) is *domain-liftable for TV distance* if the TV distance between $D_{\phi_1,\Delta,w_1,\bar{w}_1}$ and $D_{\phi_2,\Delta,w_2,\bar{w}_2}$ can be computed in time polynomial in the domain size $|\Delta|$ for any fixed pair of weighting functions $(w_1, \bar{w}_1)$ and $(w_2, \bar{w}_2)$. Similarly, a class of sentences is domain-liftable for TV distance if every pair of sentences in the class is domain-liftable for TV distance.

Example 1 *Let $\phi = \forall x \forall y : (E(x, y) \to E(y, x)) \wedge \forall x : \neg E(x, x)$. The WFOMD $D_{\phi,\Delta,w,\bar{w}}$ for any weighting functions $(w, \bar{w}) : \{E\} \to \mathbb{R}_{\geq 0}$ represents the distribution of an ER graph. Then (ϕ, ϕ) is domain-liftable for TV distance, since we can compute the TV distance of the two corresponding ER graphs $\mathcal{G}_{n,p_1}$ and $\mathcal{G}_{n,p_2}$ as $\sum_{k=0}^{n(n-1)/2} \binom{n(n-1)/2}{k} |p_1^k (1-p_1)^{n(n-1)/2-k} - p_2^k (1-p_2)^{n(n-1)/2-k}|$, which is polynomial in n.*

4. A Reduction to WFOMC

In this section, we show that the TV distance computation between two WFOMDs can be reduced to WFOMCs. Actually, the computation of other probability distribution distances

1. The variant of the counting complexity class $\#\mathsf{P}$ (Valiant, 1979) where the input is unary.

between two WFOMDs that satisfy a specific computational form can also be reduced to WFOMCs. We use the same notion of domain-liftability for a distance metric d as that for TV distance.

We will need the concept of *model-counting function* defined by (Kuzelka, 2021) to establish a connection between WFOMC and the computation of TV distance. For an interpretation I and a predicate P/k, the *cardinality* of P in I is defined as

$$N(P, I) := |\{(x_1, \ldots, x_k) \mid P(x_1, \ldots, x_k) \in I\}|,$$

i.e., the number of ground atoms for P that are true in I. Define the cardinality vector of the predicates in $\mathcal{P}$ with respect to I as $\mathbf{N}(\mathcal{P}, I) = (N(P_1, I), \ldots, N(P_m, I))$.

Definition 6 (Model-Counting Function) *Given a sentence ϕ and a domain Δ, the* model-counting function (MC-function) *is defined as*

$$\mathsf{MC}_{\phi,\Delta}(\mathbf{n}) := \left|\{I \in \mathcal{M}_{\phi,\Delta} \mid \mathbf{N}(\mathcal{P}_\phi, I) = \mathbf{n}\}\right|.$$

Define the weight of a cardinality vector $\mathbf{n} = (n_1, \ldots, n_m)$ w.r.t. $(w, \bar{w})$ as

$$C_{w,\bar{w}}(\mathbf{n}) := \prod_{i=1}^{m} w(P_i)^{n_i} \cdot \prod_{i=1}^{m} \bar{w}(P_i)^{|\Delta|^{\text{arity}(P_i)} - n_i}, \tag{2}$$

where $\text{arity}(P_i)$ is the arity of predicate P_i. We can write the weight of an interpretation I as $W_{w,\bar{w}}(I) = C_{w,\bar{w}}(\mathbf{N}(\mathcal{P}_\phi, I))$. Thus the WFOMC (Kuzelka, 2021)can be computed as $\sum_{\mathbf{n}} \mathsf{MC}_{\phi,\Delta}(\mathbf{n}) \cdot C_{w,\bar{w}}(\mathbf{n})$, where the summation is over all possible cardinality vectors. Conversely, the MC-function can be computed from the WFOMCs as shown in the following proposition.[2]

Proposition 7 (Kuzelka, 2021, Proposition 4) *Let ϕ be a sentence and Δ be a domain, if for any pair of weighting functions $(w, \bar{w})$, $\mathsf{WFOMC}(\phi, \Delta, w, \bar{w})$ can be computed in time polynomial in $|\Delta|$ and the number of bits needed to represent the weights*[3]*, then MC-function $\mathsf{MC}_{\phi,\Delta}(\mathbf{n})$ can be computed in time polynomial in $|\Delta|$.*

The MC function bridges the TV distance computation and WFOMCs through the symmetry among interpretations in the interpretation space, which has been widely exploited in the literature on lifted inference and WFOMC (Milch et al., 2008; Singla and Domingos, 2008; Van den Broeck et al., 2011; Braun and Möller, 2016). We first partition the interpretation space into a polynomial number of equivalence classes such that all interpretations in the same class have the same weight difference under the two WFOMDs. Then we can compute the distance d by summing up the contributions from these polynomially many classes, where the number of interpretations in each class can be computed by WFOMCs. In the following, we state the result for a broader class of distances, which includes TV distance as a special case.

2. This proposition is a simplified version of the result in (Kuzelka, 2021), where the result is stated in terms of the *weighted* model-counting function, a generalization of the MC-function that computes the total weight of all interpretations with a given cardinality vector.
3. This condition is a little bit stronger than the conventional domain-liftability for WFOMC (Van den Broeck et al., 2011; Van den Broeck et al., 2021). However, as far as we know, all existing domain-liftable WFOMC algorithms satisfy this condition.

Theorem 8 *Let d be a distance of probability distributions such that for any two distributions P and Q over the same measurement space S,*

$$d(P,Q) = \sum_{x \in S} f(P(x), Q(x)),$$

where f is an arbitrary computable function: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$. For a fragment $\mathcal{F}$ of first-order logic, if the WFOMC of any sentence in $\mathcal{F}$ can be computed in time polynomial in the domain size and the number of bits needed to represent the weights, then $\mathcal{F}$ is domain-liftable for the distance d.

Proof Let $\mathcal{P}$ be a predicate vocabulary, and let ϕ_1 and ϕ_2 be two sentences over $\mathcal{P} = (P_1, \ldots, P_m)$. Given two WFOMDs $D_1 = D_{\phi_1,\Delta,w_1,\bar{w}_1}$ and $D_2 = D_{\phi_2,\Delta,w_2,\bar{w}_2}$, we have $d(D_1, D_2) = \sum_{I \in \mathcal{I}_{\mathcal{P},\Delta}} f(D_1(I), D_2(I))$ and decompose it into three parts:

$$\sum_{I \in \mathcal{M}_{\phi_1 \wedge \phi_2, \Delta}} f(D_1(I), D_2(I)) + \sum_{I \in \mathcal{M}_{\neg\phi_1 \wedge \phi_2, \Delta}} f(0, D_2(I)) + \sum_{I \in \mathcal{M}_{\phi_1 \wedge \neg\phi_2, \Delta}} f(D_1(I), 0). \tag{3}$$

Let Z_1 and Z_2 be the WFOMCs of ϕ_1 and ϕ_2 respectively. Recall the Definition 5 of D_1 and D_2 as well as the Equation (2), we have for any interpretation $I \in \mathcal{I}_{\mathcal{P},\Delta}$,

$$D_1(I) = \frac{W_{w_1,\bar{w}_1}(I)}{Z_1} = \frac{W_{w_1,\bar{w}_1}(\mathbf{N}(\mathcal{P}, I))}{Z_1}, \quad D_2(I) = \frac{W_{w_2,\bar{w}_2}(I)}{Z_2} = \frac{W_{w_2,\bar{w}_2}(\mathbf{N}(\mathcal{P}, I))}{Z_2}.$$

Thus, we can rewrite each term in (3) as follows:

$$\sum_{I \in \mathcal{M}_{\phi_1 \wedge \phi_2, \Delta}} f(D_1(I), D_2(I)) = \sum_{\mathbf{n} \in \mathcal{N}} \mathsf{MC}_{\phi_1 \wedge \phi_2, \Delta}(\mathbf{n}) \cdot f\left(\frac{W_{w_1,\bar{w}_1}(\mathbf{n})}{Z_1}, \frac{W_{w_2,\bar{w}_2}(\mathbf{n})}{Z_2}\right),$$

$$\sum_{I \in \mathcal{M}_{\neg\phi_1 \wedge \phi_2, \Delta}} f(0, D_2(I)) = \sum_{\mathbf{n} \in \mathcal{N}} \mathsf{MC}_{\neg\phi_1 \wedge \phi_2, \Delta}(\mathbf{n}) \cdot f\left(0, \frac{W_{w_2,\bar{w}_2}(\mathbf{n})}{Z_2}\right),$$

$$\sum_{I \in \mathcal{M}_{\phi_1 \wedge \neg\phi_2, \Delta}} f(D_1(I), 0) = \sum_{\mathbf{n} \in \mathcal{N}} \mathsf{MC}_{\phi_1 \wedge \neg\phi_2, \Delta}(\mathbf{n}) \cdot f\left(\frac{W_{w_1,\bar{w}_1}(\mathbf{n})}{Z_1}, 0\right),$$

where $\mathcal{N} = \bigotimes_{i=1}^{m} \{0, 1, \ldots, |\Delta|^{\text{arity}(P_i)}\}$ is the set of all possible cardinality vectors. Note that $|\mathcal{N}|$ is polynomial in $|\Delta|$. By Proposition 7, the proof is completed. ■

The TV distance is a special case of the distance d in Theorem 8 with $f(x, y) = \frac{1}{2}|x - y|$. Based on the above proof of Theorem 8, we can derive Algorithm 1 that efficiently computes the TV distance between two WFOMDs by calling WFOMC as a subroutine. We use this algorithm in our experiments.

Corollary 9 *Any fragment of first-order logic that satisfies the condition in Theorem 8 is domain-liftable for TV distance.*

There are plenty of fragments and fragments with additional constraints that have been shown to satisfy the condition in Corollary 9 (Van den Broeck et al., 2011, 2014; Kuzelka, 2021; van Bremen and Kuželka, 2023; Tóth and Kuželka, 2023; Kuang et al., 2024). The direct corollary of Corollary 9 is that the TV distance between two WFOMDs or MLNs in all these fragments is also domain-liftable for TV distance.

Algorithm 1: Computing the TV distance between two WFOMDs

Input: Two WFOMDs $D_1 = D_{\phi_1,\Delta,w_1,\bar{w}_1}$ and $D_2 = D_{\phi_2,\Delta,w_2,\bar{w}_2}$
Output: The TV distance $d_{TV}(D_1, D_2)$

$Z_1 \leftarrow \mathsf{WFOMC}(\phi_1, \Delta, w_1, \bar{w}_1)$
$Z_2 \leftarrow \mathsf{WFOMC}(\phi_2, \Delta, w_2, \bar{w}_2)$
$Z_{\neg 12} \leftarrow \mathsf{WFOMC}(\neg\phi_1 \wedge \phi_2, \Delta, w_2, \bar{w}_2)$
$Z_{1\neg 2} \leftarrow \mathsf{WFOMC}(\phi_1 \wedge \neg\phi_2, \Delta, w_1, \bar{w}_1)$
$d \leftarrow Z_{\neg 12}/Z_2 + Z_{1\neg 2}/Z_1$
foreach $\mathbf{n} \in \mathcal{N}$ **do**
  $d_{\mathbf{n}} \leftarrow \left| \frac{W_{w_1,\bar{w}_1}(\mathbf{n})}{Z_1} - \frac{W_{w_2,\bar{w}_2}(\mathbf{n})}{Z_2} \right|$
  $d \leftarrow d + \mathsf{MC}_{\phi_1\wedge\phi_2,\Delta}(\mathbf{n}) \cdot d_{\mathbf{n}}$
end
return $d/2$

5. Hardness Results

In this section, we provide negative results on the computation of the TV distance between two WFOMDs. We prove the result by reducing the FOMC (Definition 2), to the TV distance computation.

Theorem 10 *Let $\mathcal{F}$ be a fragment of first-order logic that is domain-liftable for TV distance. If $\mathcal{F}$ allows arbitrary boolean combinations of the formulas in the fragment with nullary atomic formulas, then FOMC of any sentence in $\mathcal{F}$ can be computed in time polynomial in the domain size.*

Proof Let (ϕ, Δ) be an instance of the FOMC problem, where ϕ is a sentence in the fragment. We use $M_\top$ and $M_\bot$ to denote $\mathsf{FOMC}(\phi, \Delta)$ and $\mathsf{FOMC}(\neg\phi, \Delta)$, respectively. The value of $M_\top + M_\bot$ is exactly $|\mathcal{I}_{\mathcal{P}_\phi,\Delta}|$, which is denoted as M.

Define $\phi' := (\neg A() \vee \phi) \wedge (A() \vee \neg\phi)$, where $A/0$ is a new nullary predicate symbol. The new sentence ϕ' is in the fragment $\mathcal{F}$ due to the assumption that arbitrary boolean combinations of formulas in $\mathcal{F}$ with nullary atomic formulas are allowed. Let $(w_1, \bar{w}_1)$ be a pair of weighting functions for ϕ' with $w_1(A) = 2$, $\bar{w}_1(A) = 1$ and $w_1(P) = \bar{w}_1(P) = 1$ for all other predicates P. Let $(w_2, \bar{w}_2)$ be another pair of weighting functions for ϕ' with $w_2(P) = \bar{w}_2(P) = 1$ for all predicates P. Let us compute the TV distance between the two WFOMDs $D_1 = D_{\phi',\Delta,w_1,\bar{w}_1}$ and $D_2 = D_{\phi',\Delta,w_2,\bar{w}_2}$: For any interpretation $I \models \phi$, we have $D_1(I) = 2/(2M_\top + M_\bot)$ and $D_2(I) = 1/M$, and for any interpretation $I \models \neg\phi$, we have $D_1(I) = 1/(2M_\top + M_\bot)$ and $D_2(I) = 1/M$. Thus the TV distance d between D_1 and D_2 can be written as

$$d = \frac{1}{2} \cdot M_\top \cdot \left(\frac{2}{2M_\top + M_\bot} - \frac{1}{M} \right) + \frac{1}{2} \cdot M_\bot \cdot \left(\frac{1}{M} - \frac{1}{2M_\top + M_\bot} \right).$$

Solving the above equation for $M_\top$ with $M_\bot = M - M_\top$ gives us $M_\top = \frac{1}{2}M \cdot \sqrt{d^2 - 6d + 1} \pm \frac{1}{2}M(d-1)$. Assume that $d \neq 0$, or otherwise the FOMC value is trivially zero or M (it is

easy to check which is the case). Since $M_\top = \mathsf{FOMC}(\phi, \Delta)$ is a non-negative integer, the value of $M_\bot$ can only be

$$\frac{1}{2}M \cdot \sqrt{d^2 - 6d + 1} - \frac{1}{2}M \cdot (d - 1). \tag{4}$$

By the condition that the fragment is domain-liftable for TV distance, we can compute d in time polynomial in n. It can be show that the computation of $M_\top$ by Equation (4) is also polynomial in n. We can perform a binary search to find the value of $M_\top$, and as $M_\top$ is an integer and $\log M$ is polynomial in n, this will take only time polynomial in n as well. ■

FOMCs for the three-variable fragment of first-order logic ($\mathbf{FO}^3$), have been shown to be in $\#\mathsf{P}_1$-complete (Beame et al., 2015). Thus, we immediately have the following corollary.

Corollary 11 *If $\#\mathsf{P}_1 \not\subseteq \mathsf{FP}$, then $\mathbf{FO}^3$ is not domain-liftable for TV distance.*

6. Experiments

In this section we report experiments in which we used our lifted algorithm (Algorithm 1) for computing the TV distance of WFOMDs, in order to gain interesting insights into several problems, which we detail in the following. All experiments were conducted on a KVM-based virtual machine with an Intel Core (Broadwell) 60-core processor at 3.09 GHz, a X86 architecture and one NVIDIA A100 GPU.

6.1. Hard Rules Vs. Soft Rules in MLNs

Folk knowledge among practitioners who have worked with MLNs[4] is that using rules with weights greater than 10 (or a similar rather small number) is essentially the same as using hard rules. In this section we study this assertion using the lens of TV distances.

In particular, we consider two variants of a *smokers-drinkers-friends* MLN given below. One of the variants will contain hard rules whereas the other one will approximate them by soft rules. The rules used in the *smokers-drinkers-friends* MLNs are:

$$\begin{aligned}
&\alpha_1 = \neg fr(x,x), && \alpha_2 = fr(x,y) \Rightarrow fr(y,x),\\
&\alpha_3 = sm(x) \Rightarrow dr(x), && \alpha_4 = dr(x) \Rightarrow sm(x),\\
&\alpha_5 = sm(x) \wedge fr(x,y) \Rightarrow sm(y), && \alpha_6 = dr(x) \wedge fr(x,y) \Rightarrow dr(y),
\end{aligned}$$

and the two respective MLNs are:

$$\begin{aligned}
\Phi_1 =&\{(\alpha_1, \infty), (\alpha_2, \varpi), (\alpha_3, 1), (\alpha_4, 1), (\alpha_5, 2), (\alpha_6, 2)\},\\
\Phi_2 =&\{(\alpha_1, \infty), (\alpha_2, \infty), (\alpha_3, 1), (\alpha_4, 1), (\alpha_5, 2), (\alpha_6, 2)\}.
\end{aligned}$$

The predicates fr, sm, and dr represent, respectively, *friendship*, *smoking*, and *drinking*. Both Φ_1 and Φ_2 model influence from friends on the smoking and drinking habits and that there is a correlation between smoking and drinking. The only difference between the two

4. One of the authors has heard it from a prominent researcher who used to work on MLNs and who will remain anonymous.

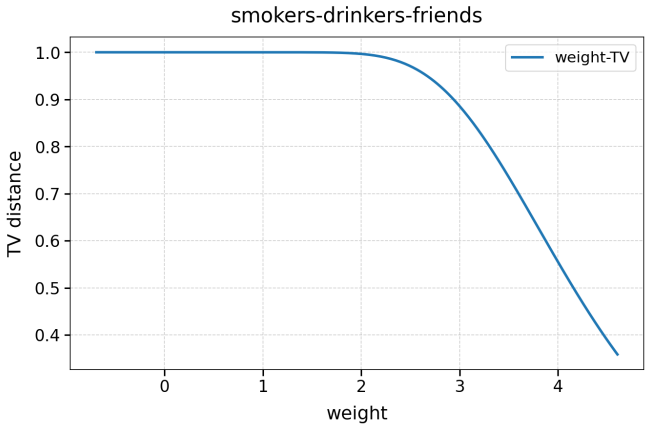


Figure 1: TV distance of "smokers-drinkers-friends" MLNs Φ_1 and Φ_2 in Section 6.1.

MLNs is that friendship in Φ_1 may be one-way, that is, there may be false friends, because the rule $fr(x,y) \Rightarrow fr(y,x)$ is not a hard rule in Φ_1, but has a finite weight ϖ. On the other hand, friendship in Φ_2 is two-way, i.e., symmetric.

To test whether the folk knowledge about approximating hard rules by soft rules is correct, we computed TV distances of Φ_1 and Φ_2 on domain of size 10 and the weight ϖ ranging from 0 to $\ln 100 \approx 4.6$, the weights were sampled with increments of 0.5 of exp(weight), which is sufficient to capture the changes in TV distance. As expected from the folk knowledge, the TV distance decreases fast after certain threshold.

6.2. Benchmarking Graph Neural Networks

We use tractable TV-distance computation between MLNs to benchmark GNN-based graph classification. GNNs perform well on graph learning tasks (Gilmer et al., 2017; Zhou et al., 2020) and are connected to first-order logic (Barceló et al., 2020; Grohe, 2024). Our procedure has three stages. First, we define two graph classes using domain-liftable MLNs Φ_1 and Φ_2 and compute the TV distance between their distributions. Next, we use exact MLN sampling (Wang et al., 2024) to generate labeled training data and train a binary GNN classifier. Finally, we evaluate it on an independently sampled held-out test set.

Since the MLNs are from a fragment domain-liftable for TV distance, we can first convert the MNLs into corresponding WFOMDs, then efficiently compute their TV distance using Algorithm 1 presented in this paper, and thus obtain the optimal Bayes error rate for classifying graphs from the two classes as $(1 - d_{TV}(\Phi_1, \Phi_2))/2$.

We consider MLNs of the form $\Phi(\varpi) = \{(\alpha_1, \infty), (\alpha_2, \infty), (\alpha_3, \varpi), (\alpha_4, \infty)\}$, where

$$\begin{aligned}
&\alpha_1 = \neg edge(x,x), \quad \alpha_2 = edge(x,y) \Rightarrow edge(y,x),\\
&\alpha_3 = edge(x,y) \Rightarrow \neg(red(x) \wedge red(y)) \wedge \neg(blue(x) \wedge blue(y)) \wedge \neg(green(x) \wedge green(y)),\\
&\alpha_4 = (red(x) \wedge \neg blue(x) \wedge \neg green(x)) \vee (\neg red(x) \wedge blue(x) \wedge \neg green(x))\\
&\qquad \vee (\neg red(x) \wedge \neg blue(x) \wedge green(x)).
\end{aligned}$$

and ϖ is a parameter. Intuitively, these MLNs model undirected graphs where each vertex has precisely one of the three colors *red*, *green*, and *blue*, and there is a soft rule α_3 with

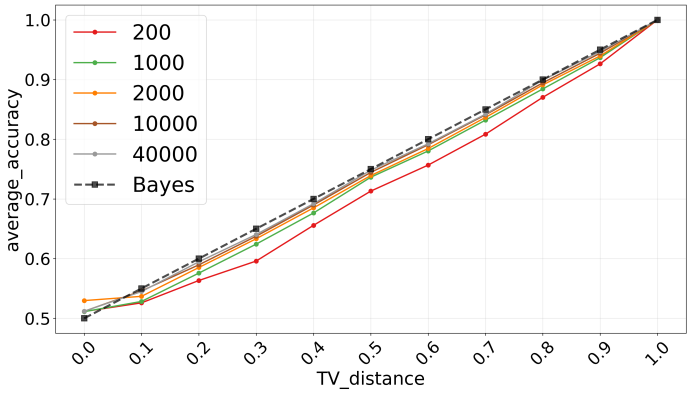


Figure 2: Accuracy of GNNs trained on datasets with varying TV distance between the distribution of the positive and negative class. These experiments with training sample sizes of 40k, 10k, 2k, 1k, and 200 were conducted.

weight ϖ that makes adjacent vertices with the same color less likely (if we set $\varpi = \infty$, the MLN would just encode properly 3-colored graphs).

We trained GNNs on datasets sampled from two MLNs with different values of ϖ, resulting in varying TV distances between the distributions. The performance of the GNNs was evaluated on another test set sampled from the same two MLNs. We varied the size of the training set to observe its impact on the accuracy of the GNNs. For each configuration of TV distance and training set size, we repeated the experiment five times and reported the average accuracy. The results are shown in Figure 2. The main observations are:

- The accuracy of the GNNs closely matches the Bayes' accuracy computed using the TV distance, indicating that GNNs can effectively learn the optimal classification boundary; it aligns with the logical expressiveness of GNNs (Barceló et al., 2020);
- There is a bend in the accuracy curve for smaller training set sizes, which might correspond to the agnostic learning phase (Haussler, 1988), where the model struggles to learn effectively due to insufficient data.

7. Conclusion

We investigate the computational complexity of the TV distance of MLNs through the lens of WFOMDs. We show both positive and negative results based on reductions between TV distance computation and WFOMCs. From the perspective of the number of logical variables in the WFOMDs, we demonstrate that the domain-liftability boundary of TV distance computation lies between the two-variable and three-variable fragments of first-order logic. We also present an application of our tractability results to benchmark GNNs on graph classification tasks. We anticipate that the new benchmarking would serve as a useful tool for evaluating future graph learning models. Finally, we note that the approximation complexity of TV distance for fragments beyond the domain-liftable ones remains an open question.

Acknowledgments

This work was supported by the Provincial Natural Science Foundation of Hunan (Grant No. 2024JJ6206), which is gratefully acknowledged by the authors. Peter Jung and Ondřej Kuželka were supported by the Czech Science Foundation project 24-11820S (Automatic Combinatorialist). Yuanhong Wang is supported by National Natural Science Foundation of China (No.62506141). Yuyi Wang is supported by the Hunan Provincial Natural Science Foundation (No. 2024JJ5128).

References

Pablo Barceló, Egor V. Kostylev, Mikaël Monet, Jorge Pérez, Juan L. Reutter, and Juan Pablo Silva. The Logical Expressiveness of Graph Neural Networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, ICLR 2020. OpenReview.net, 2020.

Paul Beame, Guy Van den Broeck, Eric Gribkoff, and Dan Suciu. Symmetric weighted first-order model counting. In *Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS '15, page 313–328, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450327572.

Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. On approximating total variation distance. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 3479–3487. International Joint Conferences on Artificial Intelligence Organization, 8 2023. doi: 10.24963/ijcai.2023/387. URL https://doi.org/10.24963/ijcai.2023/387. Main Track.

Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. Total variation distance meets probabilistic inference. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Arnab Bhattacharyya, Sutanu Gayen, Philips George John, Sayantan Sen, and N. V. Vinodchandran. Distribution learning meets graph structure sampling. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 21411–21432. Curran Associates, Inc., 2025. doi: 10.52202/085713-0722. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/1ecbfd5ea3a5745d5302c0c3ee35eb3a-Paper-Conference.pdf.

Tanya Braun and Ralf Möller. Lifted junction tree algorithm. In Gerhard Friedrich, Malte Helmert, and Franz Wotawa, editors, *KI 2016: Advances in Artificial Intelligence*, pages 30–42, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46073-4.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (2. ed.)*. Wiley, 2006.

Rodrigo De Salvo Braz, Eyal Amir, and Dan Roth. Lifted first-order probabilistic inference. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, IJCAI'05, page 1319–1325, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

Weiming Feng, Heng Guo, Mark Jerrum, and Jiaheng Wang. A simple polynomial-time approximation algorithm for the total variation distance between two product distributions. *TheoretiCS*, 2, 2023.

Weiming Feng, Liqiang Liu, and Tianren Liu. On deterministically approximating total variation distance. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 1766–1791. SIAM, 2024.

Marcel Gehrke, Tanya Braun, and Ralf Möller. Lifted dynamic junction tree algorithm. In *International Conference on Conceptual Structures*, pages 55–69. Springer, 2018.

Marcel Gehrke, Ralf Möller, and Tanya Braun. Taming reasoning in temporal probabilistic relational models. *arXiv preprint arXiv:1911.07040*, 2019.

Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning*, volume 1. MIT Press, Cambridge, 2007.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html.

Martin Grohe. The descriptive complexity of graph neural networks. *TheoretiCS*, 3, 2024.

David Haussler. Quantifying inductive bias: Ai learning algorithms and valiant's learning framework. *Artificial intelligence*, 36(2):177–221, 1988.

Kristian Kersting. Lifted Probabilistic Inference. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012*, volume 242 of *ECAI 2012*, pages 33–38, 2012. ISBN 978-1-61499-097-0.

Qipeng Kuang, Ondřej Kuželka, Yuanhong Wang, and Yuyi Wang. Bridging weighted first order model counting and graph polynomials. *arXiv preprint arXiv:2407.11877*, 2024.

Ondrej Kuzelka. Weighted first-order model counting in the two-variable fragment with counting quantifiers. *Journal of Artificial Intelligence Research*, 70:1281–1307, 2021. ISSN 1076-9757.

Ondřej Kuželka. Counting and Sampling Models in First-Order Logic. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI-2023, pages 7020–7025, Macau, SAR China, 2023. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-1-956792-03-4.

Sagar Malhotra, Davide Bizzaro, and Luciano Serafini. Lifted inference beyond first-order logic. *Artificial Intelligence*, 342:104310, 2025.

Brian Milch, Luke S. Zettlemoyer, Kristian Kersting, Michael Haimes, and Leslie Pack Kaelbling. Lifted probabilistic inference with counting formulas. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 1062–1068. AAAI Press, 2008. ISBN 9781577353683.

Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

Nima Taghipour, Daan Fierens, Guy Van den Broeck, Jesse Davis, and Hendrik Blockeel. Completeness Results for Lifted Variable Elimination. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2013, Scottsdale, AZ, USA, April 29 - May 1, 2013*, AISTATS 2013, pages 572–580, 2013.

David Poole. First-order probabilistic inference. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, page 985–991, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

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

Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press, 2014.

Parag Singla and Pedro M Domingos. Lifted first-order belief propagation. In *AAAI*, volume 8, pages 1094–1099, 2008.

N. Taghipour, D. Fierens, J. Davis, and H. Blockeel. Lifted Variable Elimination: Decoupling the Operators from the Constraint Language. *Journal of Artificial Intelligence Research*, 47:393–439, 2013. ISSN 1076-9757.

Jan Tóth and Ondřej Kuželka. Lifted inference with linear order axiom. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12295–12304, 2023.

Leslie G. Valiant. The Complexity of Enumeration and Reliability Problems. *SIAM Journal on Computing*, 8(3):410–421, 1979. ISSN 0097-5397, 1095-7111.

Timothy van Bremen and Ondřej Kuželka. Lifted inference with tree axioms. *Artificial Intelligence*, 324:103997, 2023. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2023.103997. URL https://www.sciencedirect.com/science/article/pii/S0004370223001431.

Guy Van den Broeck. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/846c260d715e5b854ffad5f70a516c88-Paper.pdf.

Guy Van den Broeck. First-order model counting in a nutshell. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 4086–4089. AAAI Press, 2016. ISBN 9781577357704.

Guy Van den Broeck, Nima Taghipour, Wannes Meert, Jesse Davis, and Luc De Raedt. Lifted probabilistic inference by first-order knowledge compilation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Three*, IJCAI'11, page 2178–2185. AAAI Press, 2011. ISBN 9781577355151.

Guy Van den Broeck, Wannes Meert, and Adnan Darwiche. Skolemization for weighted first-order model counting. In *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning*, KR'14, page 111–120. AAAI Press, 2014. ISBN 1577356578.

Guy Van den Broeck, Kristian Kersting, Sriraam Natarajan, and David Poole. *An Introduction to Lifted Probabilistic Inference*. The MIT Press, 2021. ISBN 978-0-262-36559-8.

Jan Van Haaren, Guy Van den Broeck, Wannes Meert, and Jesse Davis. Lifted generative learning of Markov logic networks. *Machine Learning*, 103(1):27–55, 2016. ISSN 15730565.

Yuanhong Wang, Juhua Pu, Yuyi Wang, and Ondřej Kuželka. Lifted algorithms for symmetric weighted first-order model sampling. *Artificial Intelligence*, 331:104114, 2024.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.