Statistical Relational AI

Exploiting Symmetries

Tanya Braun, University of Münster
Marcel Gehrke, University of Lübeck
Marco Wilhelm, TU Dortmund University

Agenda

1. Introduction [Tanya]
   - Relational models under uncertainty
     - Probabilistic Datalog, Problog, Markov logic networks (MLNs), parfactor models
     - Symmetries, semantics, inference tasks
   - Lifted inference in probabilistic relational models
     - Weighted first-order model counting and knowledge compilation in MLNs
     - Lifted variable elimination in parfactor models
2. Exploiting Symmetries in Probabilistic Graphical Models [Marcel]
3. Exploiting Symmetries in Conditional Knowledge Bases [Marco]
4. Summary [Tanya]

Statistical Relational Artificial Intelligence (StaRAI)

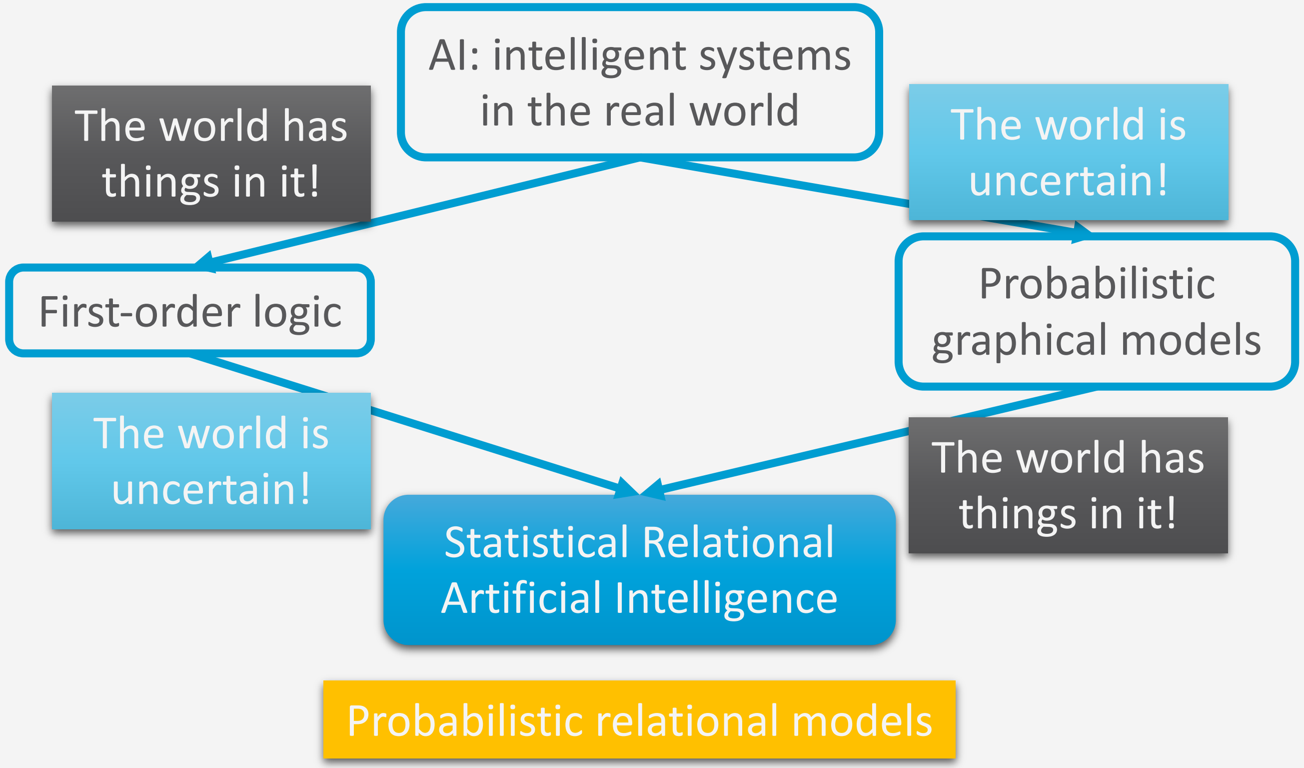


Figure based on a figure by Stuart Russell

Application

- Probabilistic Datalog for information retrieval [Fuhr 95]

```
0.7 term(d1,ir).
0.8 term(d1,db).
0.5 link(d2,d1).
about(D,T):- term(D,T).
about(D,T):- link(D,D1), about(D1,T).
```

- Query / answer

```
:- term(X,ir) & term(X,db).
X = 0.56 d1
```

Application

- Probabilistic Datalog for information retrieval [Fuhr 95]

```
0.7 term(d1,ir).
0.8 term(d1,db).
0.5 link(d2,d1).
about(D,T):- term(D,T).
about(D,T):- link(D,D1), about(D1,T).
```

- Query / answer

```
q(X):- term(X,ir).
q(X):- term(X,db).
:-q(X)
X= 0.94 d1
```

Application

- Probabilistic Datalog for information retrieval [Fuhr 95]

```
0.7 term(d1,ir).
0.8 term(d1,db).
0.5 link(d2,d1).
about(D,T):- term(D,T).
about(D,T):- link(D,D1), about(D1,T).
```

- Query / answer

```
:- about(X,db).
X= 0.8 d1;
X= 0.4 d2
```

Application

- Probabilistic Datalog for information retrieval [Fuhr 95]

```
0.7 term(d1,ir).
0.8 term(d1,db).
0.5 link(d2,d1).
about(D,T):- term(D,T).
about(D,T):- link(D,D1), about(D1,T).
```

- Query / answer

```
:- about(X,db)& about(X,ir).
X= 0.56 d1;
X= 0.28 d2  # NOT naively 0.14 = 0.8*0.5*0.7*0.5
```

Application

- Probabilistic Datalog for information retrieval [Fuhr 95]

```
0.7 term(d1,ir).
0.8 term(d1,db).
0.5 link(d2,d1).
about(D,T):- term(D,T).
about(D,T):- link(D,D1), about(D1,T).
```

- Uncertain rules?

```
0.9 temp1.
0.7 temp2.
about(D,T):- term(D,T), temp1.
about(D,T):- link(D,D1), about(D1,T), temp2.
```

ProbLog

```
% Intensional probabilistic facts:
0.6::heads(C) :- coin(C).

% Background information:
coin(c1).
coin(c2).
coin(c3).
coin(c4).

% Rules:
someHeads :- heads(_).

% Queries:
query(someHeads).
0.9744
```

https://dtai.cs.kuleuven.be/problog/

ProbLog

- Probabilistic inference [De Raedt et al. 07]
  - Compute marginal probabilities of any number of ground atoms in the presence of evidence
  - Sample from a ProbLog program
    - Generate random structures
      - Use case: [Goodman & Tenenbaum 16]
- Learn the parameters of a ProbLog program from partial interpretations [Fierens et al. 15]

Markov Logic Networks (MLNs)

- Weighted logical formulas to soften otherwise hard constraints [Richardson & Domingos 06]
  - Implicitly connected via conjunction
    - I.e., set of formulas ψ_i = knowledge base/theory
  - Worlds that violate constraint become less likely but not impossible
    - As w_i increases, so does the strength of ψ_i
    - Infinite weight: Hard constraint = pure logic formula
      - Probabilities of worlds that do not satisfy hard constraint set to 0

Hard constraint

$\infty\ Presents(X, P, C) \Rightarrow Attends(X, C)$

Soft constraint, weight = $\exp(3.75)$

$3.75\ Publishes(X, C) \wedge FarAway(C) \Rightarrow Attends(X, C)$

Grounding

- Each (w_i, ψ_i) represents a set of *propositional* sentences, each sentence with weight w_i
  - One sentence for each possible substitution of the free variables $free(\psi_i)$ in ψ_i given a finite domain (or a constraint set) D over $free(\psi_i)$
    - $\theta_D = \bigcup_{\boldsymbol{d} \in D} \{\bigcup_{t \in \boldsymbol{d}} \{X_d \to t\}\}$
  - Example: MLN $\Psi = \{(w_i, \psi_i)\}_{i=1}^{2}$
    - Domains
      - $dom(X) = \{alice\}$
      - $dom(P) = \{p_1, p_2\}$
      - $dom(C) = \{ijcai, kr\}$
    - Groundings on the right

$$\big(10, Presents(alice, p_1, ijcai) \Rightarrow Attends(alice, ijcai)\big)$$
$$\big(10, Presents(alice, p_1, kr) \Rightarrow Attends(alice, kr)\big)$$
$$\big(10, Presents(alice, p_2, ijcai) \Rightarrow Attends(alice, ijcai)\big)$$
$$\big(10, Presents(alice, p_2, kr) \Rightarrow Attends(alice, kr)\big)$$
$$\big(3.75, Publishes(alice, ijcai) \wedge FarAway(ijcai) \Rightarrow Attends(alice, ijcai)\big)$$
$$\big(3.75, Publishes(alice, kr) \wedge FarAway(kr) \Rightarrow Attends(alice, kr)\big)$$

$10 \; Presents(X, P, C) \Rightarrow Attends(X, C)$

$3.75 \; Publishes(X, C) \wedge FarAway(C) \Rightarrow Attends(X, C)$

MLNs: Semantics

- MLN $\Psi = \{(w_i, \psi_i)\}_{i=1}^{n}$, with $w_i \in \mathbb{R}$, induces a *probability distribution* over all possible interpretations ω (world) of the grounded atoms in Ψ

$$\omega \in \{true, false\}^N$$

  - N = the number of ground atoms in the grounded Ψ
  - Probability of one interpretation ω

$$P(\omega) = \frac{1}{Z} \prod_{i=1}^{n} \exp\left(w_i \cdot n_i(\omega)\right) = \frac{1}{Z} \exp\left(\sum_{i=1}^{n} w_i n_i(\omega)\right)$$

    - $n_i(\omega)$ = number of propositional sentences of ψ_i that evaluate to $true$ given the assignments of ω

MLN: Graphical Representation?

- Usually not depicted by a graph but by the logical formulas with their weights to the left
- Since the name invokes Markov networks, which is a graphical model, let us build an analogue:
  - Logical atoms as nodes
  - Edges between atoms whenever atoms occur together in a formula
    - Each ψ_i forms clique in graph
    - Potential function ϕ_i for each clique from weights using $\exp w_i$ for each model and $\exp 0$ otherwise

$Presents(X,P,C)$	$Attends(X,C)$	ϕ_1
$false$	$false$	$\exp 10$
$false$	$true$	$\exp 10$
$true$	$false$	$\exp 0$
$true$	$true$	$\exp 10$

$10\ Presents(X,P,C) \Rightarrow Attends(X,C)$

$3.75\ Publishes(X,C) \land FarAway(C) \Rightarrow Attends(X,C)$

From Weighted Formulas to Parfactors

- MLNs with their logical formulas have the same value w_i for each interpretation the satisfies ψ_i
- Allowing for different values for each interpretation, i.e., arbitrary distributions in potential functions

→ Set of parfactors
  - Parfactor: Factor (potential function) whose arguments are parameterised with logical variables
  - An MLN can be translated into a set of parfactors and vice versa [Van den Broeck 13]

$Presents(X,P,C)$	$Attends(X,C)$	ϕ_1
$false$	$false$	$\exp 10$
$false$	$true$	$\exp 10$
$true$	$false$	$\exp 0$
$true$	$true$	$\exp 10$

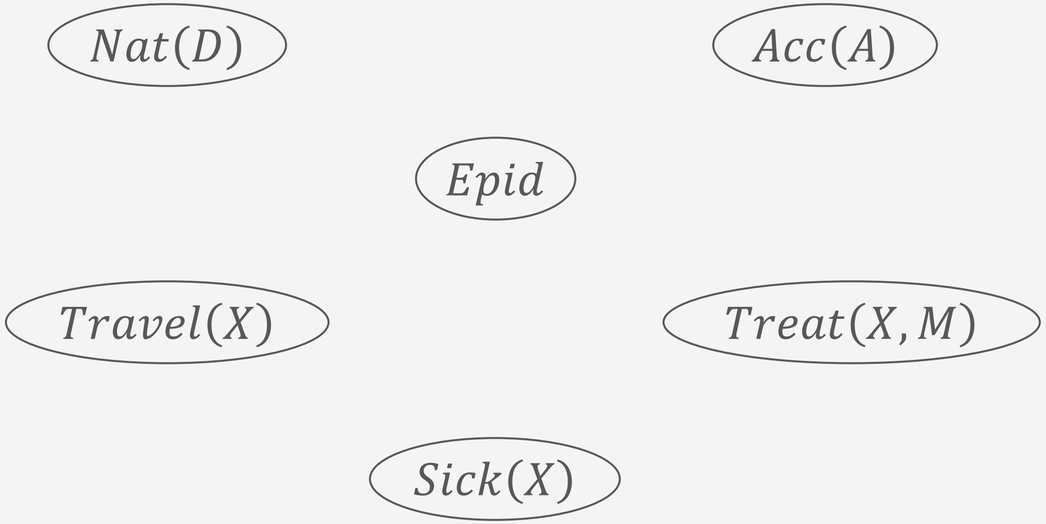
$10\ Presents(X,P,C) \Rightarrow Attends(X,C)$

$3.75\ Publishes(X,C) \wedge FarAway(C) \Rightarrow Attends(X,C)$

Logical Variables in Random Variables

- Atoms: Parameterised random variables = PRVs
  - With logical variables
    - E.g., X, M
    - Possible values (domain):
      $$dom(X) = \{alice, eve, bob\}$$
      $$dom(M) = \{injection, tablet\}$$
  - With range
    - E.g., Boolean, but any discrete, finite set possible
    - $ran(Travel(X)) = \{true, false\}$
  - Represent sets of *indistinguishable* random variables

$$Nat(D) = natural\ disaster\ D$$
$$Acc(A) = accident\ A$$

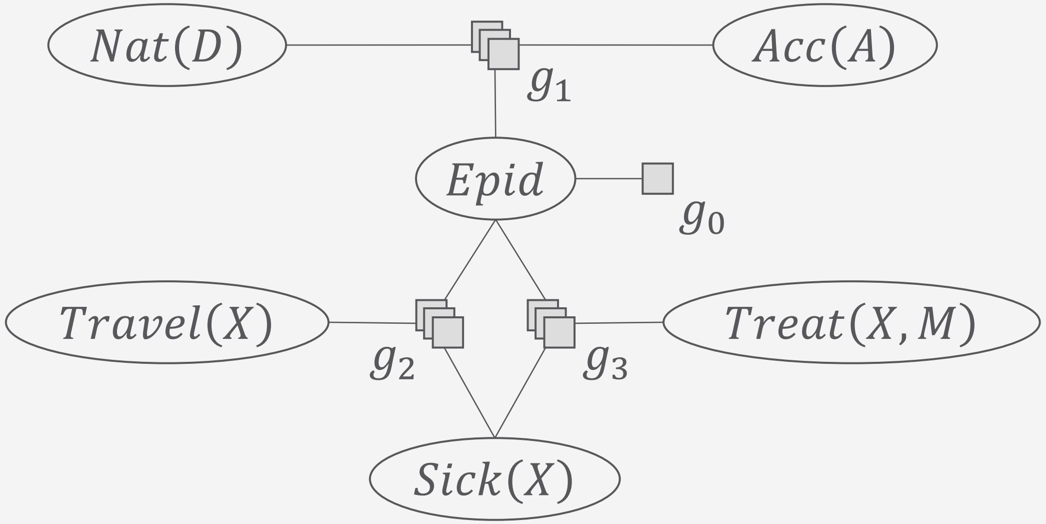


Parfactors

- Factors with PRVs = parfactors
  - E.g., g_2

$Travel(X)$	$Epid$	$Sick(X)$	g_2
$false$	$false$	$false$	5
$false$	$false$	$true$	0
$false$	$true$	$false$	4
$false$	$true$	$true$	6
$true$	$false$	$false$	4
$true$	$false$	$true$	6
$true$	$true$	$false$	2
$true$	$true$	$true$	9

Potentials
- In parfactors, just like in factors, no probability distribution as factors required



Factors

- Grounding
  - E.g., $gr(g_2) = \{f_2^1, f_2^2, f_2^3\}$

$Travel(X)$	$Epid$	$Sick(X)$	g_2
$false$	$false$	$false$	5
$false$	$false$	$true$	0
$false$	$true$	$false$	4
$false$	$true$	$true$	6
$true$	$false$	$false$	4
$true$	$false$	$true$	6
$true$	$true$	$false$	2
$true$	$true$	$true$	9

$Travel(eve)$	$Epid$	$Sick(eve)$	g_2
$false$	$false$	$false$	5
$false$	$false$	$true$	0
$false$	$true$	$false$	4
$false$	$true$	$true$	6
$true$	$false$	$false$	4
$true$	$fals$		
$true$	tru		
$true$	tru		

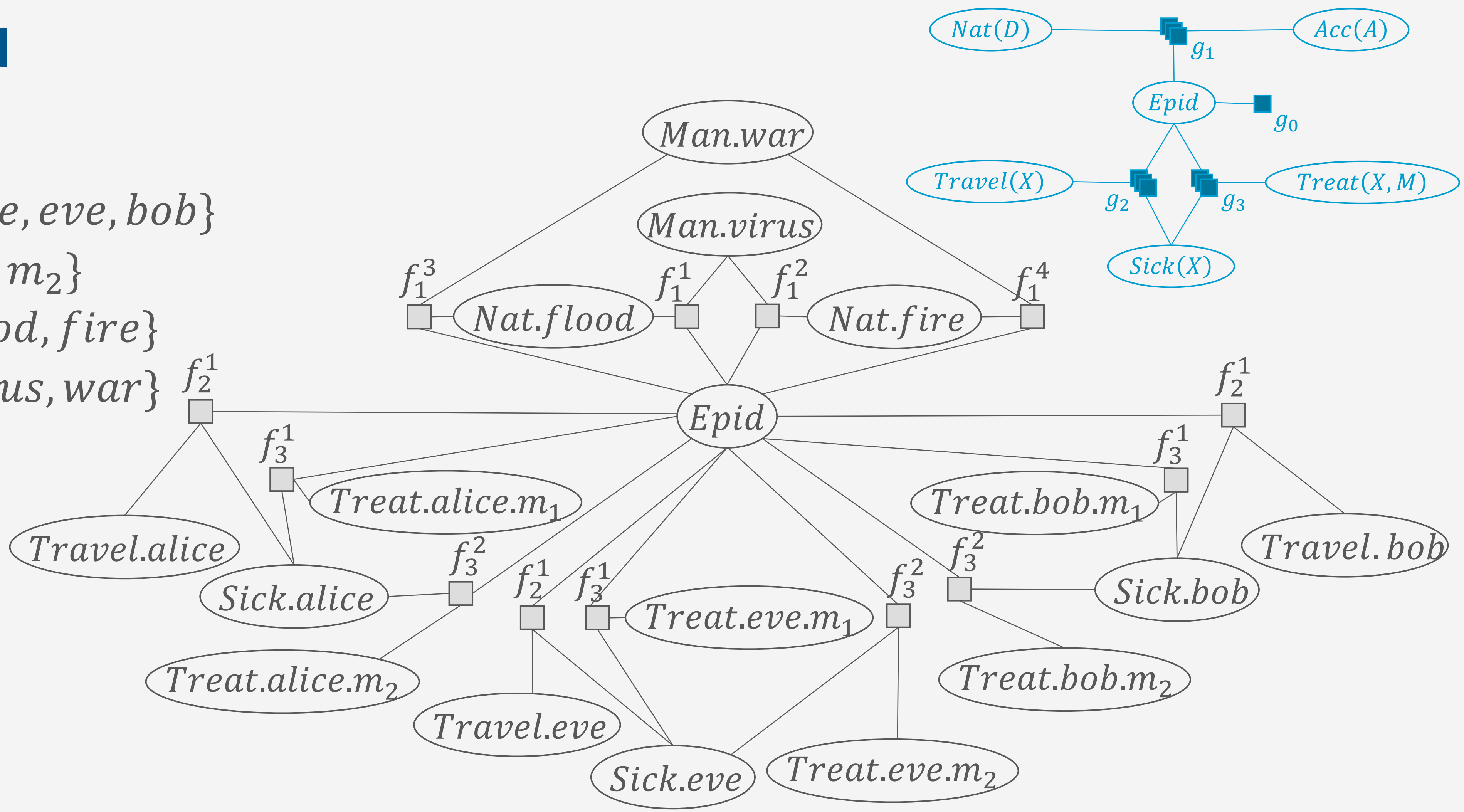
$Travel(bob)$	$Epid$	$Sick(bob)$	g_2
$false$	$false$	$false$	5
$false$	$false$	$true$	0
$false$	$true$	$false$	4
		$true$	6
		$false$	4
		$true$	6
		$false$	2
		$true$	9

$Travel(alice)$	$Epid$	$Sick(alice)$	g_2
$false$	$false$	$false$	5
$false$	$false$	$true$	0
$false$	$true$	$false$	4
$false$	$true$	$true$	6
$true$	$false$	$false$	4
$true$	$false$	$true$	6
$true$	$true$	$false$	2
$true$	$true$	$true$	9

$reat(X, M)$

Grounded Model

- Given domains
  - $dom(X) = \{alice, eve, bob\}$
  - $dom(M) = \{m_1, m_2\}$
  - $dom(D) = \{flood, fire\}$
  - $dom(W) = \{virus, war\}$
- Symmetry in
  - Graph structure (isomorphism)
  - Factors (identity)



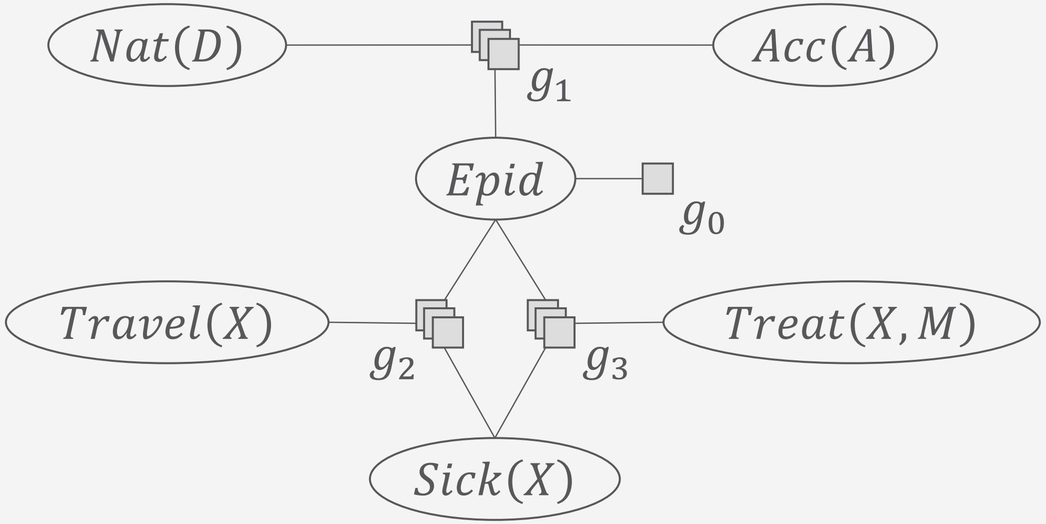
Encoding the Joint Distribution

- Set of parfactors = model
  - E.g., $G = \{g_1, g_2, g_3\}$
  - Semantics: Joint probability distribution P_G
    - Build by grounding, multiplying all grounded factors, and normalising the result
    - Grounding semantics [Sato 95, Fuhr 95]

$$P_G = \frac{1}{Z} \prod_{f \in gr(G)} f$$

$$Z = \sum_{v \in r(rv(gr(G)))} \prod_{f \in gr(G)} f_i(\pi_{rv(f_i)}(v))$$

$\pi_{variables}(v)$ = projection of v onto $variables$



Semantics

- Distribution semantics (aka grounding or Herbrand semantics) [Sato 95]
  - Completely define discrete joint distribution by factorisation
  - Probabilistic extensions to Datalog [Fuhr 95]
  - Relational Bayesian networks [Jaeger 97]
  - Bayesian Logic Programming [Milch et al. 05], ProbLog [De Raedt et al. 07]
  - Parfactor models [Poole 03, Taghipour et al. 13, B & Möller 18, Gehrke et al. 19]
  - Markov logic networks (MLNs) [Richardson & Domingos 06]
- Probabilistic Soft Logic (PSL) [Bach et al. 17]
  - Define density function using log-linear model
- Maximum entropy semantics [Thimm et al. 10]
  - Partial specification of discrete joint with “uniform completion” [→ Marco]

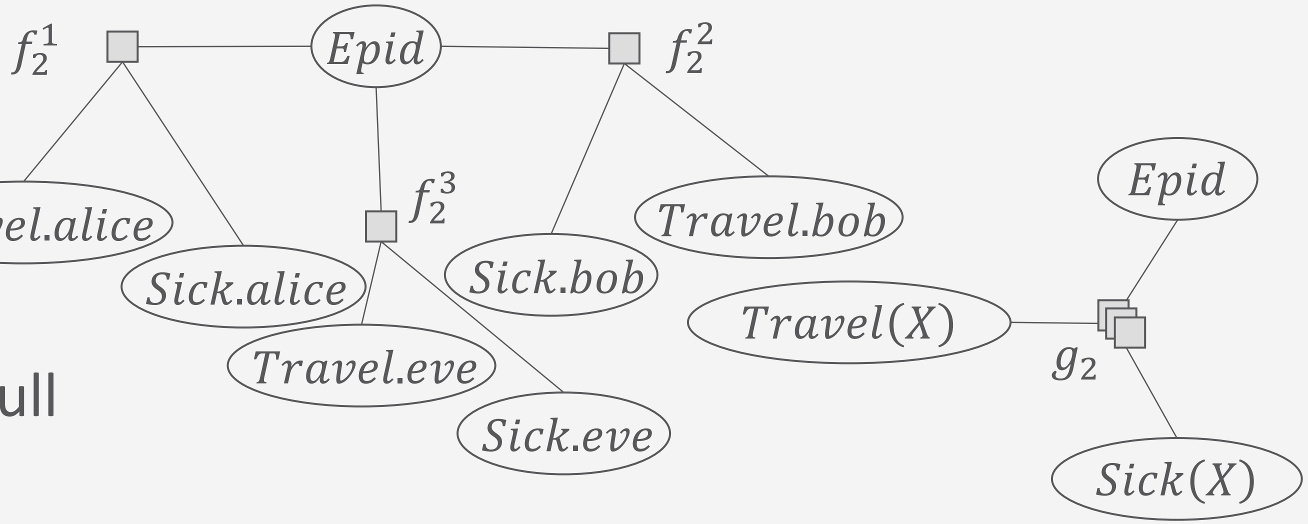
About Symmetries

- In a grounded model versus the first-order version, a form of symmetry appears
  - Logic perspective:
    - First-order relations that are identical for each grounding
  - Graph perspective:
    - Isomorphic subgraphs
  - Parameter perspective:
    - Identical weights / potentials

→ Exchangeable random variables in the full joint probability distribution

$$(10, Presents(alice, p_1, ijcai) \Rightarrow Attends(alice, ijcai))$$
$$(10, Presents(alice, p_1, kr) \Rightarrow Attends(alice, kr))$$
$$(10, Presents(alice, p_2, ijcai) \Rightarrow Attends(alice, ijcai))$$
$$(10, Presents(alice, p_2, kr) \Rightarrow Attends(alice, kr))$$

$$10\ Presents(X, P, C) \Rightarrow Attends(X, C)$$

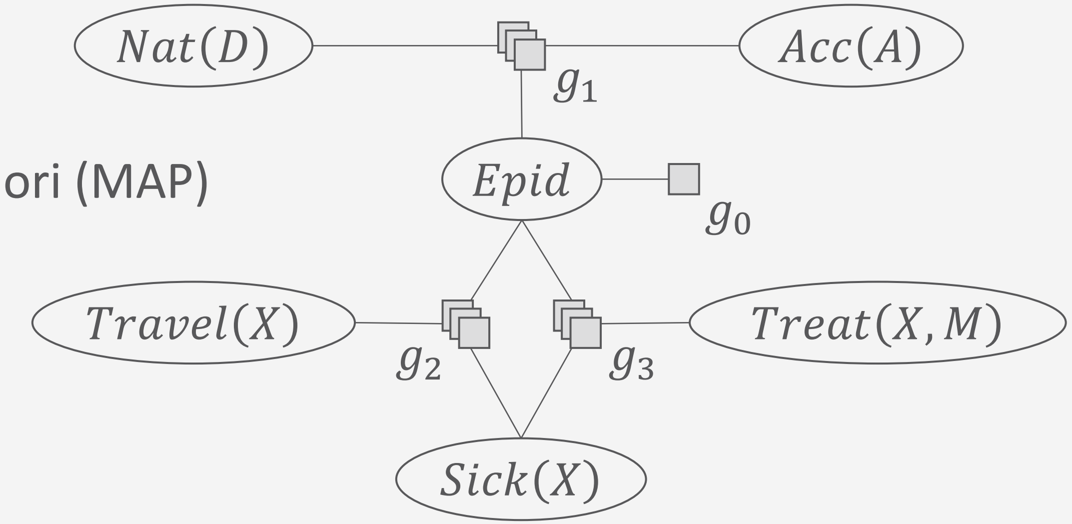


Inference Problems with and without Evidence

$10\ Presents(X,P,C) \Rightarrow Attends(X,C)$

$3.75\ Publishes(X,C) \land FarAway(C) \Rightarrow Attends(X,C)$

- Query answering problem given a model:
  - Probability of events
    - E.g., $P(Att(eve,kr) = true), P(Epid = true)$
  - Conditional (marginal) probability distributions
    - E.g., $P(Att(ev,kr)|FarAway(kr)), P(Epid|sick(alice), sick(eve))$
  - Assignment queries:
    - Most probable states of random variables
    - Most-probable explanation (MPE), Maximum a posteriori (MAP)
- Lifted inference:
  Work with representatives for exchangeable random variables
  - Avoid grounding for as long as possible



Agenda

1. Introduction [Tanya]
   - Relational models under uncertainty
     - Probabilistic Datalog, Problog, Markov logic networks (MLNs), parfactor models
     - Symmetries, semantics, inference tasks
   - Lifted inference in probabilistic relational models
     - Weighted first-order model counting and knowledge compilation in MLNs
     - Lifted variable elimination in parfactor models
2. Exploiting Symmetries in Probabilistic Graphical Models [Marcel]
3. Exploiting Symmetries in Conditional Knowledge Bases [Marco]
4. Summary [Tanya]

Weighted First-order Model Counting

- Define a weighted first-order model counting problem using a weighted first-order model count (WFOMC) [Van den Broeck et al. 11]

$$WFOMC(\Delta, w_T, w_F) = \sum_{\substack{\omega = \omega_T \cup \omega_F \\ \omega \in \Omega_\Delta}} \prod_{l \in \omega_T} w_T(pred(l)) \prod_{l \in \omega_F} w_F(pred(l))$$

  - Δ a theory in FOL with domain constraints (finite domains for logical variables)
  - w_T a weight function for predicates being positive
  - w_F a weight function for predicates being negative
  - Ω_Δ the set of worlds (i.e., models in logics) of Δ
  - $pred(l)$ a function mapping a literal l to its predicate
- Query can be answered by computing

$$P(q_i|e) = \frac{WFOMC(\Delta \wedge e \wedge q_i, w_T, w_F)}{WFOMC(\Delta \wedge e, w_T, w_F)}$$

Example

- Theory: one sentence
  $\forall X \in \text{People} : smokes(X) \Rightarrow cancer(X)$
  - People $= \{x_1, x_2\}$
  - Weight functions
    - $w_T(smokes(X)) = 3$
    - $w_F(\neg smokes(X)) = 1$
    - $w_T(cancer(X)) = 6$
    - $w_F(\neg cancer(X)) = 2$
  - Model count: 9

$$WFOMC(\Delta, w_T, w_F) = \sum_{\substack{\omega = \omega_T \cup \omega_F \\ \omega \in \Omega_\Delta}} \prod_{l \in \omega_T} w_T(pred(l)) \prod_{l \in \omega_F} w_F(pred(l))$$

$s(x_1)$	$c(x_1)$	$s(x_2)$	$c(x_2)$	Weight	
0	0	0	0	$1 \cdot 2 \cdot 1 \cdot 2$	4
0	0	0	1	$1 \cdot 2 \cdot 1 \cdot 6$	12
~~0~~	~~0~~	~~1~~	~~0~~	~~$1 \cdot 2 \cdot 3 \cdot 2$~~	~~12~~
0	0	1	1	$1 \cdot 2 \cdot 3 \cdot 6$	36
0	1	0	0	$1 \cdot 6 \cdot 1 \cdot 2$	12
0	1	0	1	$1 \cdot 6 \cdot 1 \cdot 6$	36
~~0~~	~~1~~	~~1~~	~~0~~	~~$1 \cdot 6 \cdot 3 \cdot 2$~~	~~36~~
0	1	1	1	$1 \cdot 6 \cdot 3 \cdot 6$	108
~~1~~	~~0~~	~~0~~	~~0~~	~~$3 \cdot 2 \cdot 1 \cdot 2$~~	~~12~~
~~1~~	~~0~~	~~0~~	~~1~~	~~$3 \cdot 2 \cdot 1 \cdot 6$~~	~~36~~
~~1~~	~~0~~	~~1~~	~~0~~	~~$3 \cdot 2 \cdot 3 \cdot 2$~~	~~36~~
~~1~~	~~0~~	~~1~~	~~1~~	~~$3 \cdot 2 \cdot 3 \cdot 6$~~	~~108~~
1	1	0	0	$3 \cdot 6 \cdot 1 \cdot 2$	36
1	1	0	1	$3 \cdot 6 \cdot 1 \cdot 6$	108
~~1~~	~~1~~	~~1~~	~~0~~	~~$3 \cdot 6 \cdot 3 \cdot 2$~~	~~108~~
1	1	1	1	$3 \cdot 6 \cdot 3 \cdot 6$	324
				+	676

Example

- Theory: one sentence
  $\forall X \in \text{People} : smokes(X) \Rightarrow cancer(X)$
  - People $= \{x_1, x_2\}$
  - Weight functions
    - $w_T(smokes(X)) = 3$
    - $w_F(\neg smokes(X)) = 1$
    - $w_T(cancer(X)) = 6$
    - $w_F(\neg cancer(X)) = 2$
  - Model count: 9

$$P(s(x_1)) = \frac{WFOMC(\Delta \wedge s(x_1), w_T, w_F)}{WFOMC(\Delta, w_T, w_F)} = \frac{36 + 108 + 324}{676} = \frac{468}{676}$$

$s(x_1)$	$c(x_1)$	$s(x_2)$	$c(x_2)$	Weight	
~~0~~	~~0~~	~~0~~	~~0~~	~~$1 \cdot 2 \cdot 1 \cdot 2$~~	~~4~~
~~0~~	~~0~~	~~0~~	~~1~~	~~$1 \cdot 2 \cdot 1 \cdot 6$~~	~~12~~
~~0~~	~~0~~	~~1~~	~~0~~	~~$1 \cdot 2 \cdot 3 \cdot 2$~~	~~12~~
~~0~~	~~0~~	~~1~~	~~1~~	~~$1 \cdot 2 \cdot 3 \cdot 6$~~	~~36~~
~~0~~	~~1~~	~~0~~	~~0~~	~~$1 \cdot 6 \cdot 1 \cdot 2$~~	~~12~~
~~0~~	~~1~~	~~0~~	~~1~~	~~$1 \cdot 6 \cdot 1 \cdot 6$~~	~~36~~
~~0~~	~~1~~	~~1~~	~~0~~	~~$1 \cdot 6 \cdot 3 \cdot 2$~~	~~36~~
~~0~~	~~1~~	~~1~~	~~1~~	~~$1 \cdot 6 \cdot 3 \cdot 6$~~	~~108~~
~~1~~	~~0~~	~~0~~	~~0~~	~~$3 \cdot 2 \cdot 1 \cdot 2$~~	~~12~~
~~1~~	~~0~~	~~0~~	~~1~~	~~$3 \cdot 2 \cdot 1 \cdot 6$~~	~~36~~
~~1~~	~~0~~	~~1~~	~~0~~	~~$3 \cdot 2 \cdot 3 \cdot 2$~~	~~36~~
~~1~~	~~0~~	~~1~~	~~1~~	~~$3 \cdot 2 \cdot 3 \cdot 6$~~	~~108~~
1	1	0	0	$3 \cdot 6 \cdot 1 \cdot 2$	36
1	1	0	1	$3 \cdot 6 \cdot 1 \cdot 6$	108
~~1~~	~~1~~	~~1~~	~~0~~	~~$3 \cdot 6 \cdot 3 \cdot 2$~~	~~108~~
1	1	1	1	$3 \cdot 6 \cdot 3 \cdot 6$	324
				+	676

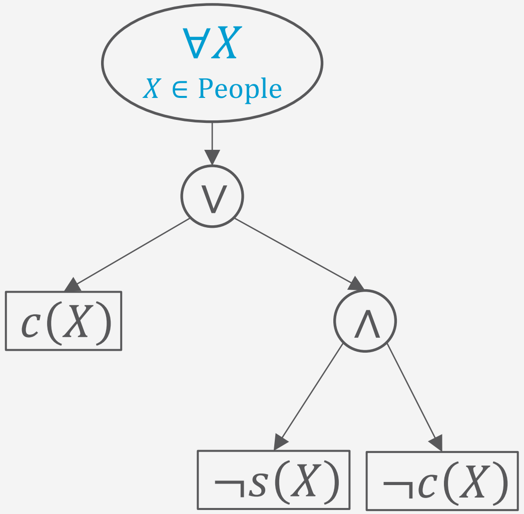
ction

First-order (FO) Circuits

- Assume theory in Skolem normal form + CNF
  - Sequence of intensional conjunctions in CNF
  - E.g., with $s = smokes, c = cancer$

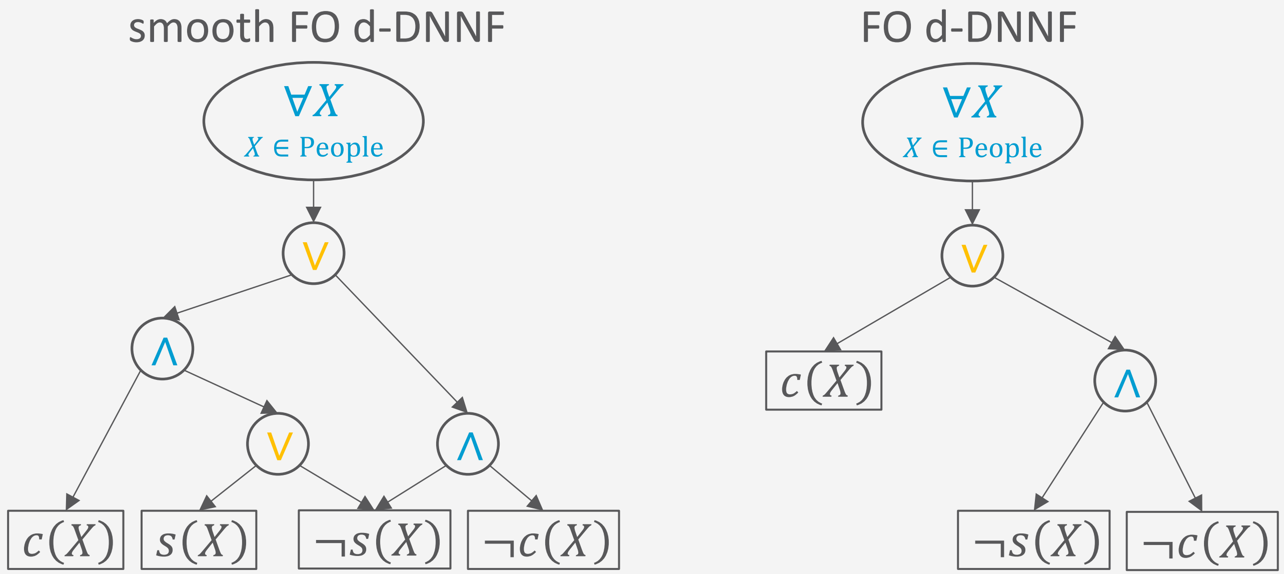
$$\forall X \in \text{People} \; : s(X) \Rightarrow c(X)$$
$$\equiv \forall X \in \text{People} \; : \neg s(X) \lor c(X)$$

- FO circuit (excerpt)
  - Inner nodes:
    - Extensional conjunctions/disjunctions (as before)
    - Set conjunctions
  - Leaf nodes
    - Positive and negative predicates, $true$, $false$
  - Full + construction: see PhD thesis by Guy Van den Broeck



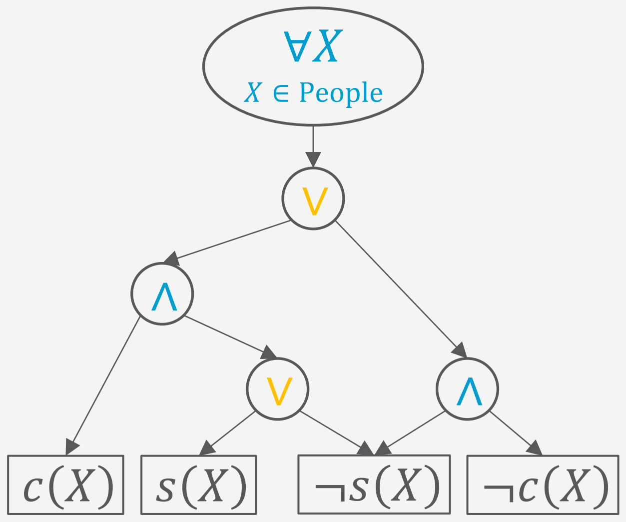
Smooth FO d-DNNF Circuits

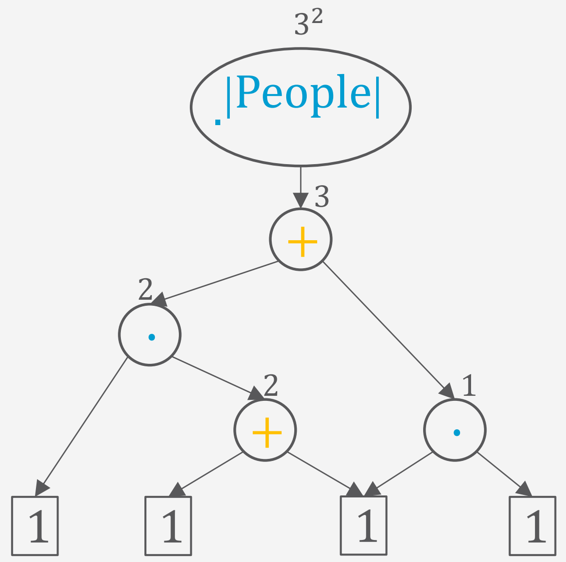
- Properties
  - Deterministic disjunctions
    - Only one disjunct (child node) can be true at the same time
  - Decomposable conjunctions
    - Each pair of conjuncts (child nodes) must be independent
  - Smoothness
    - Each disjunct contains the same variables



Arithmetic FO d-DNNF Circuits

- Given FO d-DNNF, the following replacements are possible to compute WFOMCs
  - Replace $\wedge$ with $\cdot$
  - Replace $\vee$ with $+$
  - Replace $\forall$ with exponentiation for $|\mathrm{Domain}|$
  - Replace leaves with 1's
  - E.g., with $|\mathrm{People}| = |\{x_1, x_2\}| = 2$



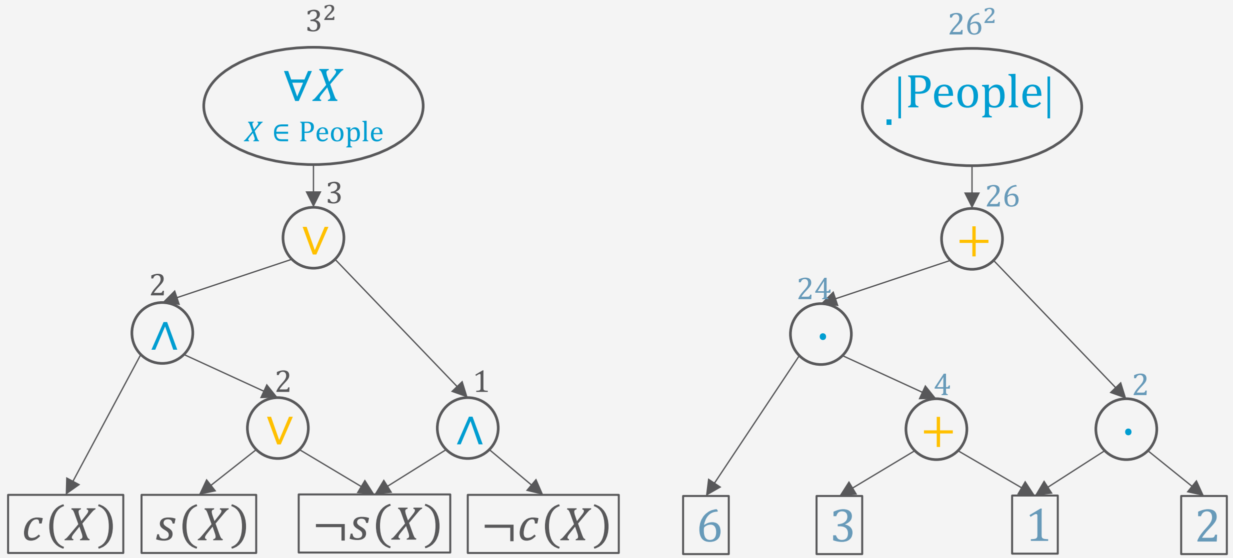


WFOMC Circuits

- Replace
  - Replace ∧ with ·
  - Replace ∨ with +
  - Replace ∀ with exponentiation for $|\text{Domain}|$
  - Replace leaves with weights
  - E.g., with $|\text{People}| = |\{x_1, x_2\}| = 2$

$$w_T(smokes(X)) = 3$$
$$w_F(\neg smokes(X)) = 1$$
$$w_T(cancer(X)) = 6$$
$$w_F(\neg cancer(X)) = 2$$

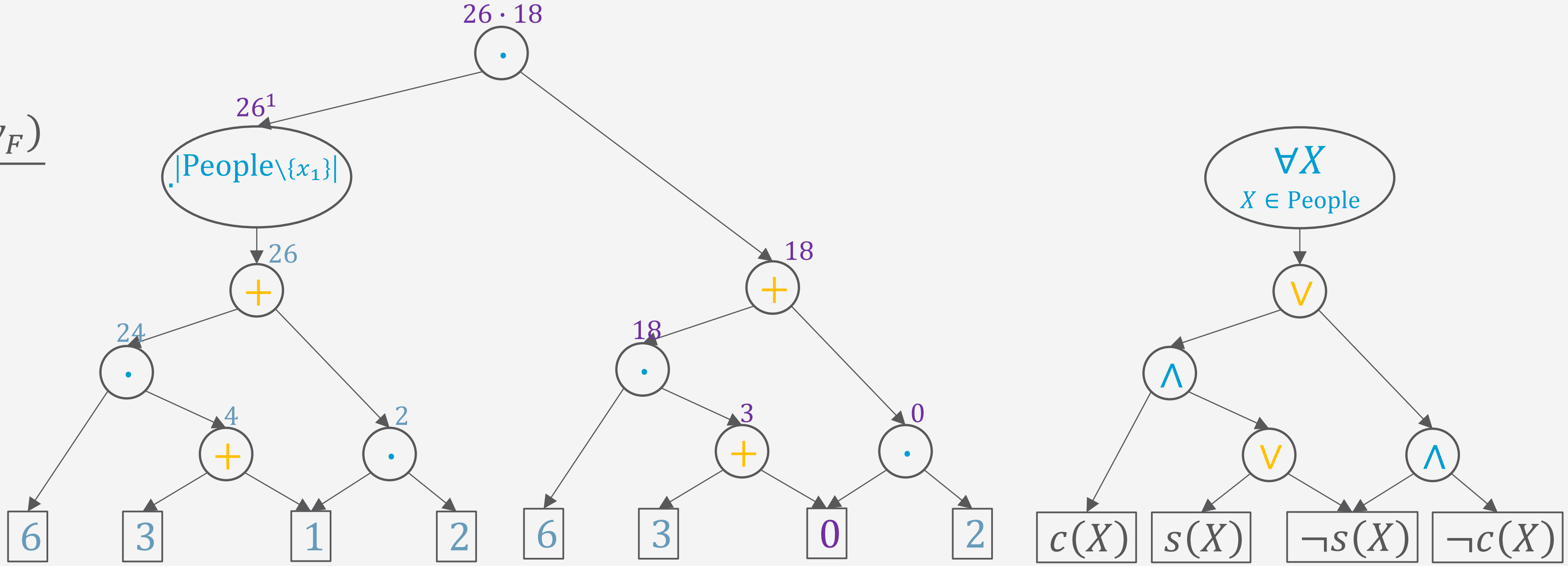
$$WFOMC(\Delta, w_T, w_F) = \sum_{\substack{\omega=\omega_T \cup \omega_F \\ \omega \in \Omega_\Delta}} \prod_{l \in \omega_T} w_T(pred(l)) \prod_{l \in \omega_F} w_F(pred(l))$$



WFOMC Circuits

- Given $P(q_i)$
  - Basically, compile a circuit for $\Delta \wedge q_i$ reusing components from the circuit of Δ
  - E.g., $P\big(s(x_1)\big)$ with $|\text{People}| = |\{x_1, x_2\}| = 2$

$$P\big(s(x_1)\big) = \frac{WFOMC(\Delta \wedge s(x_1), w_T, w_F)}{WFOMC(\Delta, w_T, w_F)} = \frac{468}{676} = 0.692$$



MLNs for WFOMCs

- Weights in MLNs specified for formulas instead of single predicates
- Trick:
  1. Introduce a new predicate θ_i containing all free variables of ψ_i as equivalent to ψ_i
     - $\forall X \in \text{People}, P \in \text{Paper}, C \in \text{Conference} \;:\; \theta_1(X,P,C) \Leftrightarrow \big(Presents(X,P,C) \Rightarrow Attends(X,C)\big)$
     - $\forall X \in \text{People}, C \in \text{Conference} : \theta_2(X) \Leftrightarrow \big(Publishes(X,C) \wedge FarAway(C) \Rightarrow Attends(X,C)\big)$
  2. Specify weight functions such that θ_i takes the weight of ψ_i
     - $w_T\big(\theta_1(X)\big) = \exp(10)$
     - $w_T\big(\theta_2(X)\big) = \exp(3.5)$
     - All other predicates, $\neg\theta_1, \neg\theta_2$ are mapped to 1 by both w_T, w_F

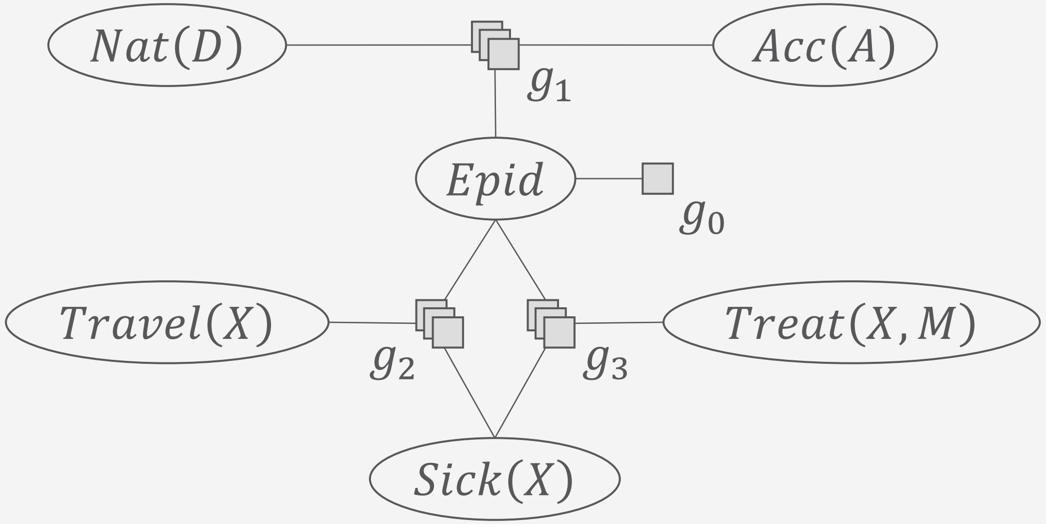
$10 \; Presents(X,P,C) \Rightarrow Attends(X,C)$

Alternative lifted solution approach for MLNs:
Probabilistic Theorem Proving
[Gogate & Domingos 11]

$3.75 \; Publishes(X,C) \wedge FarAway(C) \Rightarrow Attends(X,C)$

QA in Parfactor Models: Lifted Variable Elimination (LVE)

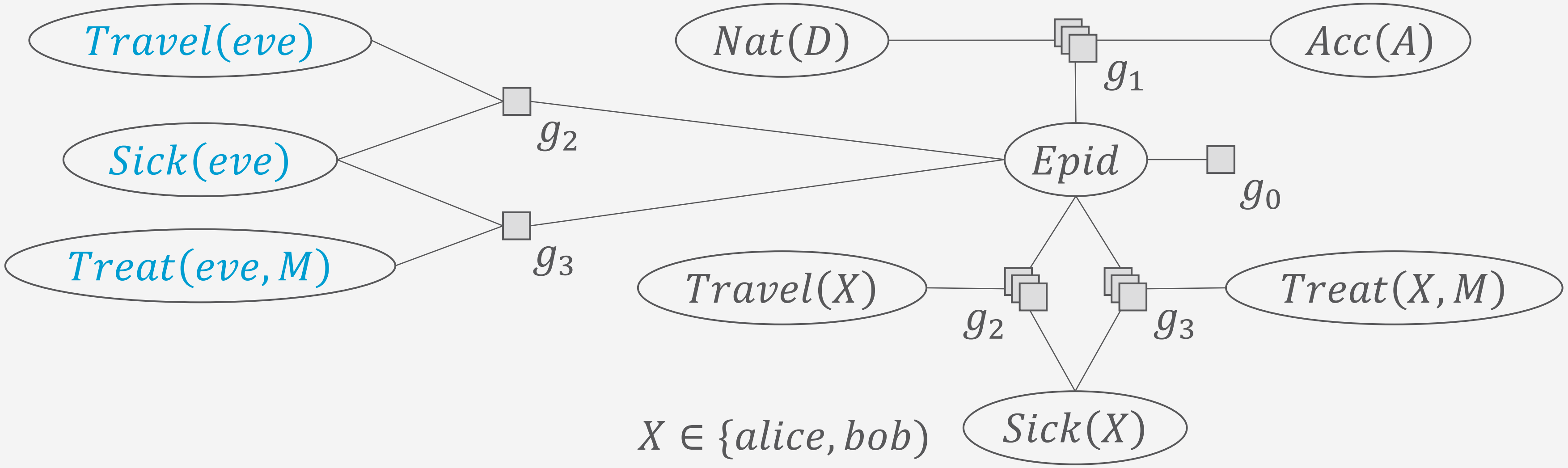
- Eliminate all variables not appearing in query
  - [Poole 03, de Salvo Braz et al. 05, 06, Milch et al. 08, Taghipour et al. 13, 13a, B & Möller 18]
- Lifted summing out
  - Sum out *representative* instance as in propositional variable elimination
  - Exponentiate result for exchangeable instances
- Correctness: Equivalent ground operation
  - Each instance is summed out
  - Result: factor f that is identical for all instance
  - Multiplying indistinguishable results
    → exponentiation of one representative f



QA: LVE in Detail

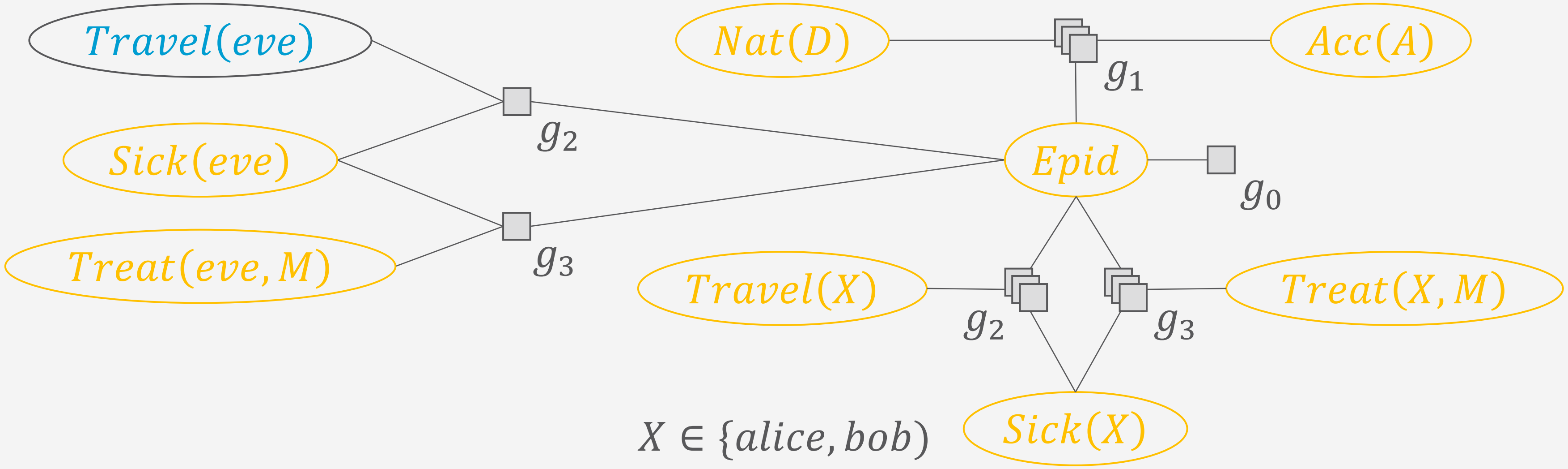
- E.g., marginal
  - $P(Travel(eve))$
  - Split atoms $R(\ldots, X, \ldots)$ w.r.t. eve if eve in $dom(X)$

Shattering



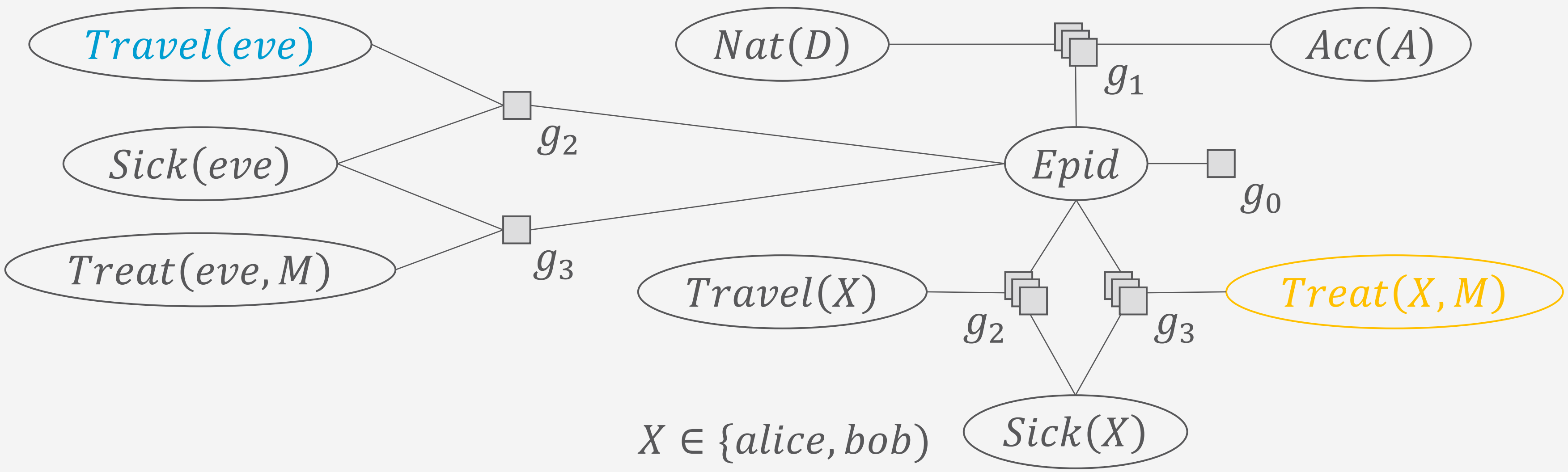
QA: LVE in Detail

- E.g., marginal
  - $P(Travel(eve))$
  - Split atoms $R(\ldots, X, \ldots)$ w.r.t. eve if eve in $dom(X)$
  - Eliminate all non-query variables



QA: LVE in Detail

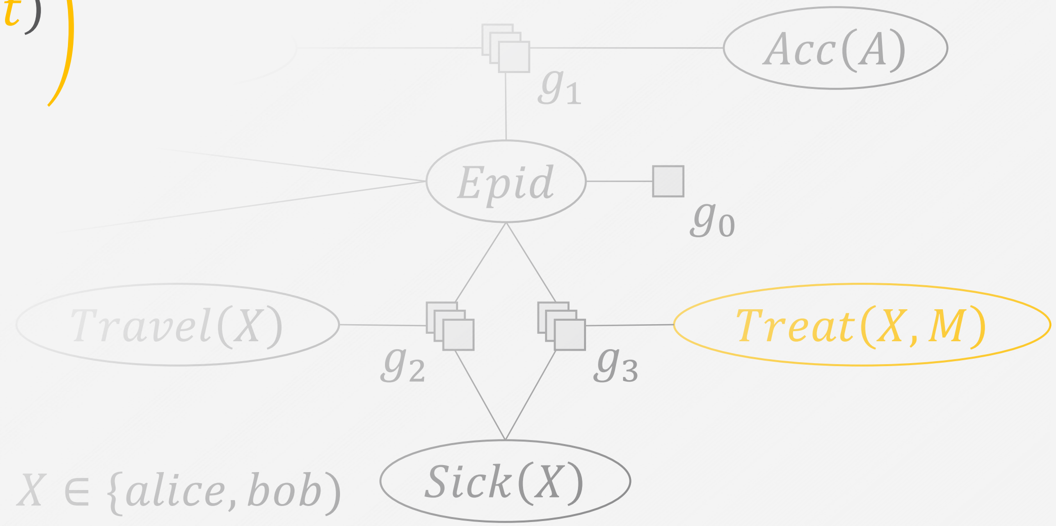
- Eliminate $Treat(X, M)$
  - Appears in only one g: g_3
  - Contains all logical variables of g_3: X, M
  - For each X constant: the same number of M constants
  - ✓ Preconditions of lifted summing out fulfilled, lifted summing out possible



LVE in Detail: Lifted Summing Out

- Eliminate $Treat(X, M)$ by lifted summing out
  1. Sum out representative
  2. Exponentiate for indistinguishable objects

$$\left(\sum_{t \in r(Treat(X,M))} g_3(Epid = e, Sick(X) = s, Treat(X,M) = t)\right)^{\#M|X}$$



LVE in Detail: Lifted Summing Out

$$\left(\sum_{t \in r(Treat(X,M))} g_3(Epid = e, Sick(X) = s, Treat(X,M) = t)\right)^{\#M|X}$$

$Epid$	$Sick(X)$	$Treat(X,M)$	g_3
$false$	$false$	$false$	9
$false$	$false$	$true$	1
$false$	$true$	$false$	6
$false$	$true$	$true$	3
$true$	$false$	$false$	7
$true$	$false$	$true$	5
$true$	$true$	$false$	4
$true$	$true$	$true$	8

+

$Epid$	$Sick(X)$	Σ
$false$	$false$	10
$false$	$true$	9
$true$	$false$	12
$true$	$true$	12

^

$Epid$	$Sick(X)$	^
$false$	$false$	10^2
$false$	$true$	9^2
$true$	$false$	12^2
$true$	$true$	12^2

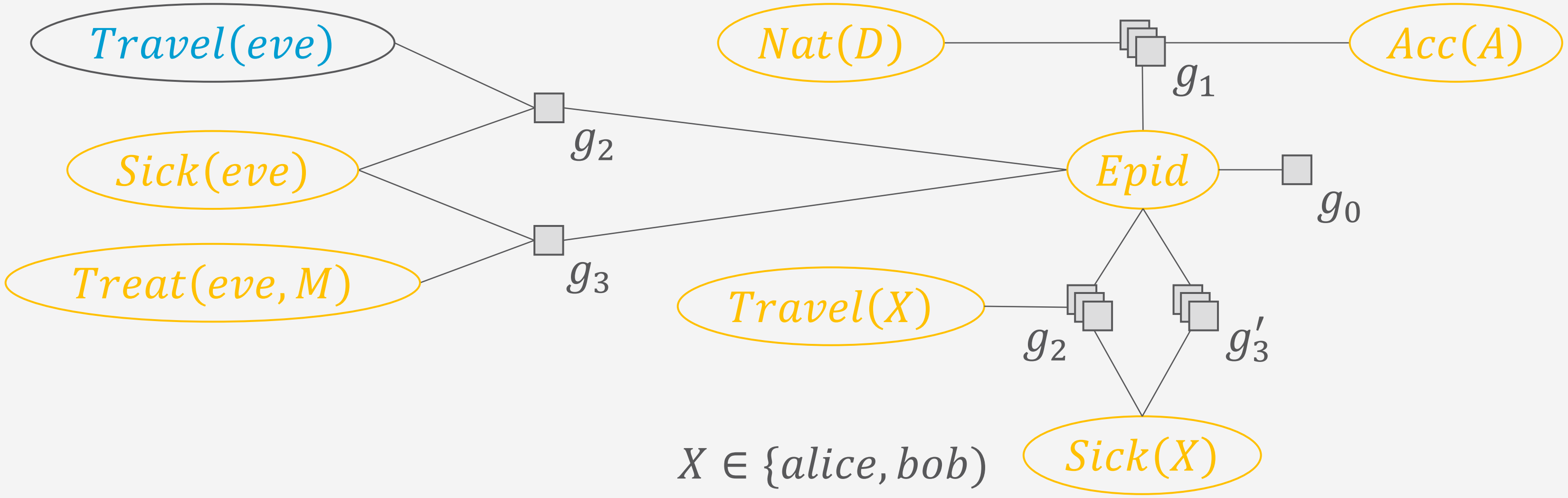
LVE in Detail: Lifted Summing Out

- Result after summing out $Treat(X, M)$

$$\left(\sum_{t \in r(Treat(X,M))} g_3(Epid = e, Sick(X) = s, Treat(X,M) = t)\right)^{\#M|X}$$

Only here, domain size comes into play
→ no change in graph / parfactor if domain size changes

$Epid$	$Sick(X)$	g_3'
$false$	$false$	100
$false$	$true$	81
$true$	$false$	144
$true$	$true$	144



Tractability

Can also be considered for evidence and query terms [Van den Broeck 13, B 20]

- Given a model that allows for lifted calculations
  - I.e., no groundings during solving an instance of the problem
- Solving an instance of the problem is possible in time polynomial in domain sizes

  → The query answering algorithm is domain-lifted
    - Domain-lifted algorithms: LVE [Taghipour et al. 13], First-order knowledge compilation [Van den Broeck 11], Lifted Junction Tree Algorithm [B 20]
- An query answering problem is tractable
  - If solved by an efficient algorithm running in time polynomial in the number of random variables
- Assume that number of random variables characterised by domain sizes
  - Then, solving a query answering problem is tractable under domain-liftability
    - Runtime might still be exponential in other terms
    - More general results by Niepert & Van den Broeck (2014)

Interim Summary

- Relational models under uncertainty
  - Probabilistic Datalog: Set of probabilistic facts and rules
  - ProbLog: Set of intensional probabilistic facts and rules
  - MLNs: Set of weighted formulas
  - Parfactor models: Set of potential functions over PRVs
  - Grounding semantics, symmetries in grounded models
- Lifted inference in probabilistic relational models
  - Query answering through WFOMC, compilation into a first-order circuit for efficient inference
  - Lifted variable elimination by summing out representatives
  - Tractability in terms of domain sizes through lifting

WWU MÜNSTER

Agenda

1. Introduction [Tanya]
2. Exploiting Symmetries in Probabilistic Graphical Models [Marcel]
   - Identifying Most Likely Sources of Events
   - Introducing Time
   - Approximating Symmetries over Time to Keep Reasoning Polynomial
3. Exploiting Symmetries in Conditional Knowledge Bases [Marco]
4. Summary [Tanya]

WWU MÜNSTER

Bibliography

Ordered alphabetically

Bibliography

- Bach et al. (2017)
  Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor: Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. In *Machine Learning Journal*, 2017.
- B & Möller (2018)
  Tanya Braun and Ralf Möller: Parameterised Queries and Lifted Query Answering. In *IJCAI-18 Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018.
- B (2020)
  Tanya Braun: Rescued from a Sea of Queries: Exact Inference in Probabilistic Relational Models. *PhD thesis*, 2020.
- De Raedt et al. (2007)
  Luc De Raedt, Angelika Kimmig, and Hannu Toivonen: ProbLog – A Probabilistic Prolog and its Application in Link Discovery, In *IJCAI-07 Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007.
- De Salvo Braz et al. (2005)
  Rodrigo de Salvo Braz, Eyal Amir, and Dan Roth: Lifted First-order Probabilistic Inference. In *IJCAI-05 Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 2005.

Bibliography

- De Salvo Braz et al. (2006)
  Rodrigo de Salvo Braz, Eyal Amir, and Dan Roth: MPE and Partial Inversion in Lifted Probabilistic Variable Elimination. In *AAAI-06 Proceedings of the 21st Conference on Artificial Intelligence*, 2006.
- Fierens et al. (2015)
  Daan Fierens, Guy Van den Broeck, Joris Renkens, Dimitar Shterionov, Bernd Gutmann, Ingo Thon, Gerda Janssens, and Luc De Raedt: Inference and learning in probabilistic logic programs using weighted Boolean formulas. In *Theory and Practice of Logic Programming*, 2015
- Fuhr (1995)
  Norbert Fuhr: Probabilistic Datalog - A Logic for Powerful Retrieval Methods. In *SIGIR-95 Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1995.
- Gehrke et al. (2019)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Relational Forward Backward Algorithm for Multiple Queries. In *FLAIRS-32 Proceedings of the 32nd International Florida Artificial Intelligence Research Society Conference*, 2019.

Bibliography

- Gogate & Domingos (2011)
  Vibhav Gogate and Pedro Domingos: Probabilistic Theorem Proving. In *UAI-11 Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, 2011.
- Jaeger (1997)
  Manfred Jaeger: Relational Bayesian Networks. In *UAI-97 Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, 1997.
- Milch et al. (2005)
  Brian Milch, Bhaskara Marthi, Stuart Russell, David Sontag, Daniel L. Long, and Andrey Kolobov: BLOG: Probabilistic Models with Unknown Objects. In *IJCAI-05 Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 2005.
- Milch et al. (2008)
  Brian Milch, Luke S. Zettelmoyer, Kristian Kersting, Michael Haimes, and Leslie Pack Kaelbling: Lifted Probabilistic Inference with Counting Formulas. In *AAAI-08 Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 2008.

Bibliography

- Niepert & Van den Broeck (2014)
  Mathias Niepert and Guy Van den Broeck. Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference. In *AAAI-14 Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014.
- Poole (2003)
  David Poole: First-order Probabilistic Inference. In *IJCAI-03 Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003.
- Richardson & Domingos (2006)
  Matthew Richardson and Pedro Domingos. Markov Logic Networks. In *Machine Learning Journal*, 2006.
- Sato (1995)
  Taisuke Sato: A Statistical Learning Method for Logic Programs with Distribution Semantics. In *Proceedings of the 12th International Conference on Logic Programming*, 1995.

Bibliography

- Taghipour et al. (2013)
  Nima Taghipour, Daan Fierens, Guy Van den Broeck, Jesse Davis, and Hendrik Blockeel. Completeness Results for Lifted Variable Elimination. In *AISTATS-13 Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, 2013.
- Taghipour et al. (2013a)
  Nima Taghipour, Daan Fierens, Jesse Davis, and Hendrik Blockeel. Lifted Variable Elimination: Decoupling the Operators from the Constraint Language. *Journal of Artificial Intelligence Research*, 47(1):393–439, 2013.
- Thimm et al. (2010)
  Matthias Thimm, Marc Finthammer, Sebastian Loh, Gebriele Kern-Isberner, and Christoph Beierle: A System for Relational Probabilistic Reasoning on Maximum Entropy. In *FLAIRS-23 Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference*, 2010.
- Van den Broeck et al. (2011)
  Guy Van den Broeck, Nima Taghipour, Wannes Meert, Jesse Davis, and Luc De Raedt: Lifted Probabilistic Inference by First-order Knowledge Compilation. In *IJCAI-11 Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011.

Bibliography

- Van den Broeck (2011)
  Guy Van den Broeck: On the Completeness of First-Order Knowledge Compilation for Lifted Probabilistic Inference. In *NIPS-11 Advances in Neural Information Processing Systems 24*, 2011.
- Van den Broeck (2013)
  Guy Van den Broeck: Lifted Inference and Learning in Statistical Relational Models. *PhD thesis*, 2013.