# COMPRESSION VS. ACCURACY:
## COMPACT MODELS FOR EFFICIENCY AND INTERPRETATION

MALTE LUTTERMANN[1], JAN SPELLER[2], MARCEL GEHRKE[1], AND TANYA BRAUN[2]

UHH

[1]Institute of Humanities-Centered Artificial Intelligence, University of Hamburg
[2]Computer Science Department, University of Münster

Universität Münster

# AGENDA

1. Introduction and background [Tanya]
   - Motivation, use cases
   - Probabilistic (relational) models
   - Lifted probabilistic inference
2. Compressing models [Malte]
3. Accuracy considerations [Jan]
4. Summary and future directions [Tanya]

# GENERAL AGENT SETTING

- Architecture
  - Entity with corresponding physical sensors and actuators
- Program
  - Implementation of an agent function
    - Utility function as a description of desirable states
      → Maximise expected utility
      → Rational agent behaviour
  - Runs on an architecture
- Both depend on task environment

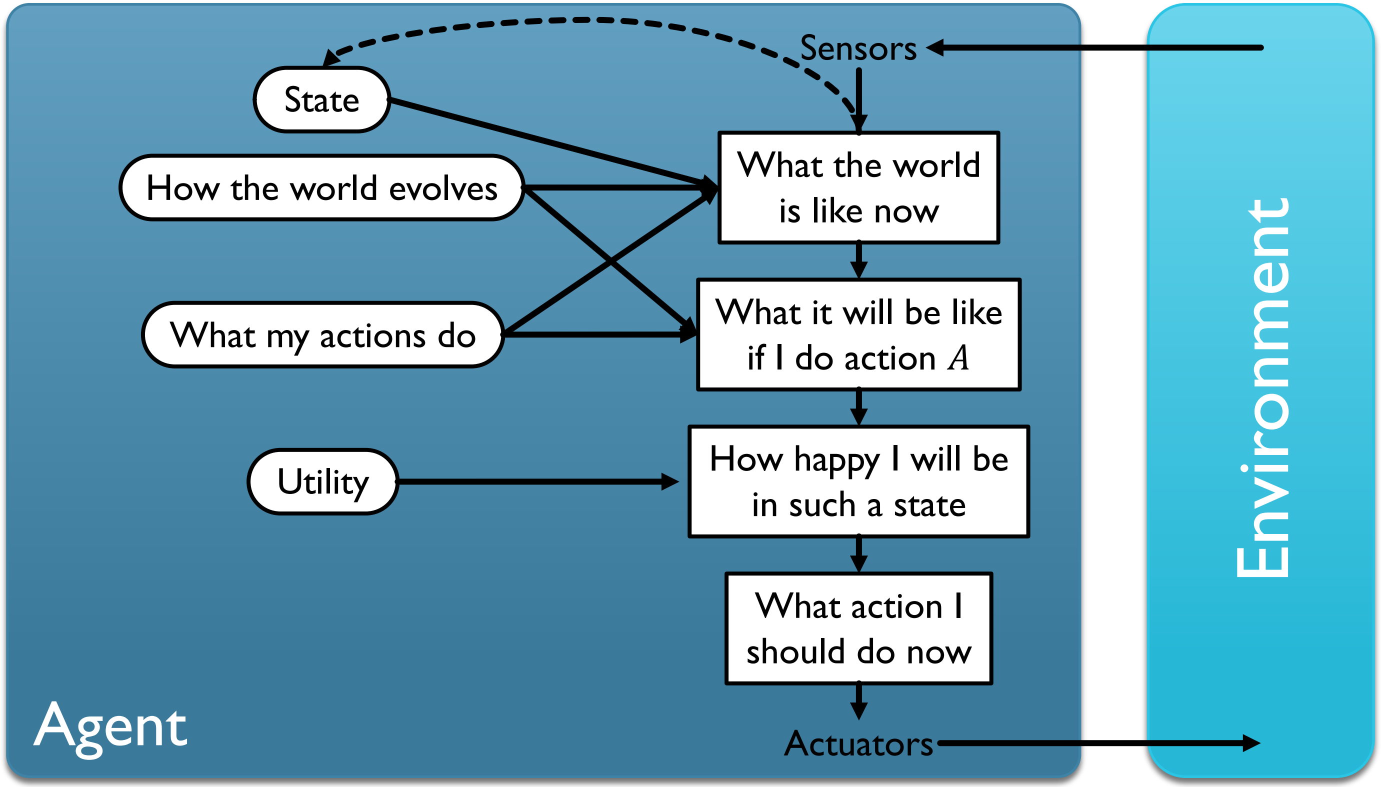



Figure based on AIMA, Russell/Norvig

# HOW DO WE REPRESENT THE STATE?

## COMPACTLY AND ACCURATELY

# DESCRIPTION OF THE ENVIRONMENT

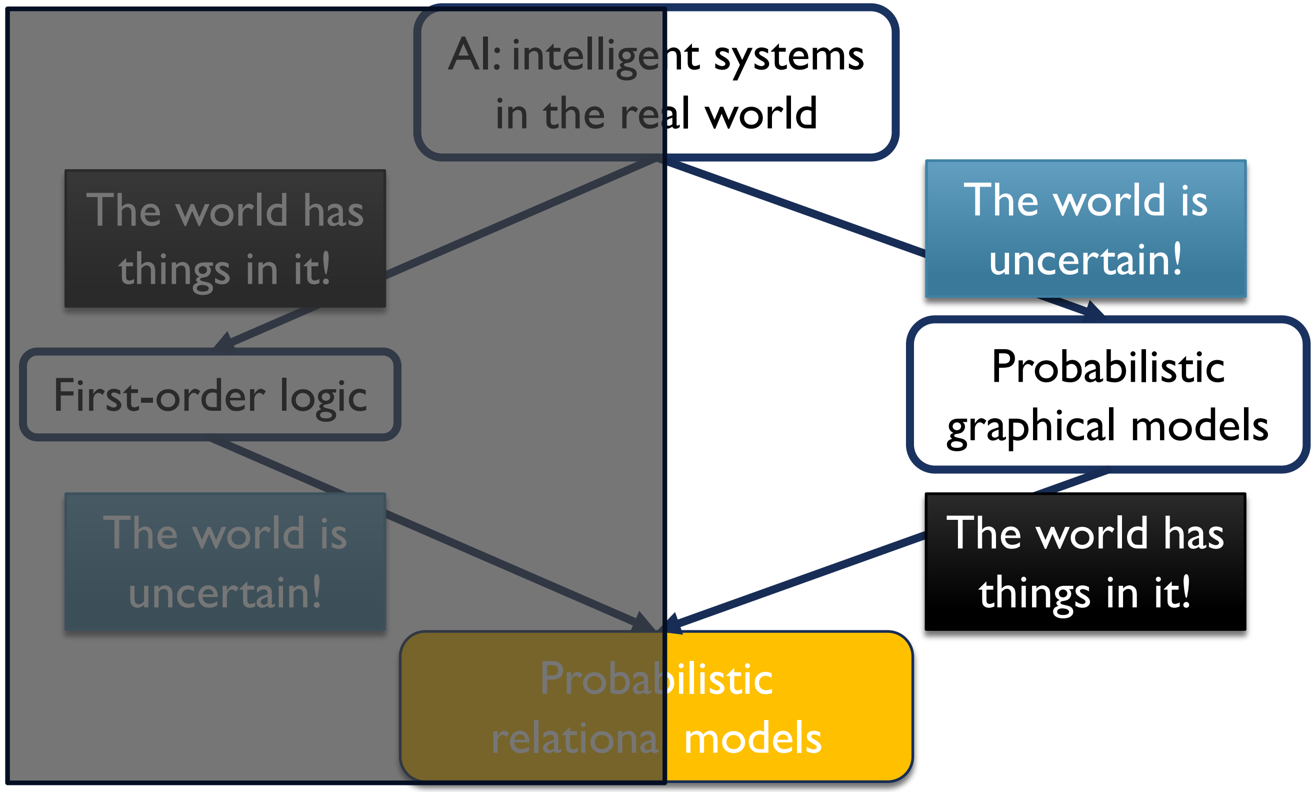



Figure based on a figure by Stuart Russell

# THE WORLD IS UNCERTAIN: PROBABILITY DISTRIBUTION

- Random variables to encode features of the environment
- Probability distribution over possible worlds
- Probabilistic reasoning
  - E.g., what is the probability of there being an epidemic?
  - Sum over all random variables not occurring in the query
- Size and runtime *exponential* in the number of random variables: $O(r^n)$
  - $n$ number of random variables
  - $r$ largest number of feature values among all random variables

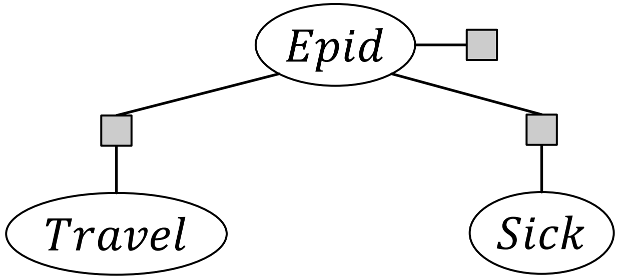
| $Epid$ | $P$ |
|---|---|
| false | 0.80 |
| true | 0.20 |

| $Epid$ | $Travel$ | $Sick$ | $P$ |
|---|---|---|---|
| false | false | false | 0.20 |
| false | false | true | 0.24 |
| false | true | false | 0.28 |
| false | true | true | 0.08 |
| true | false | false | 0.05 |
| true | false | true | 0.06 |
| true | true | false | 0.07 |
| true | true | true | 0.02 |

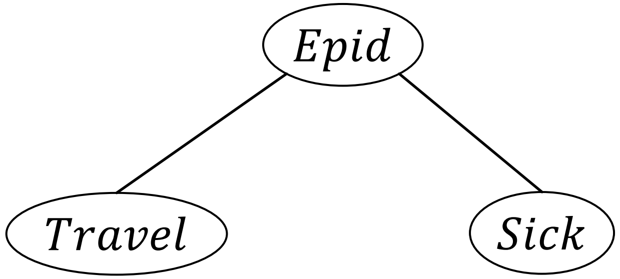
Full joint probability distribution

# PROBABILISTIC GRAPHICAL MODELS (PGMS)

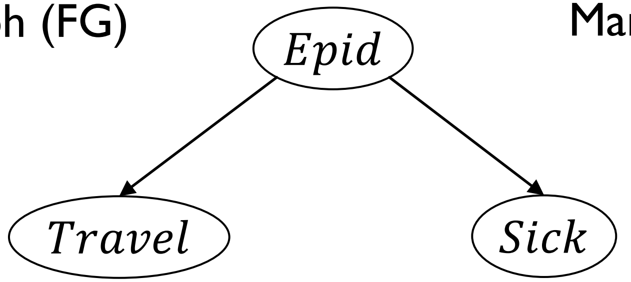
- *Factorisation* of a full joint according to (conditional) independences in the full joint




Factor graph (FG)




Markov network (MN)



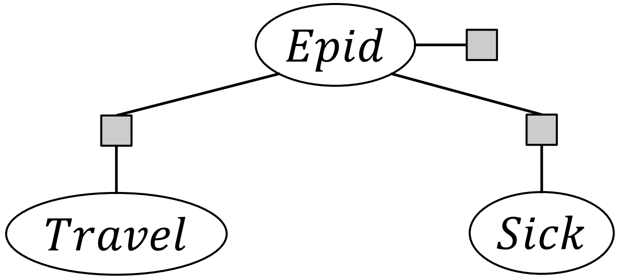

Bayesian network (BN)

| *Epid* | *Travel* | *Sick* | *P* |
|---|---|---|---|
| false | false | false | 0.20 |
| false | false | true | 0.24 |
| false | true | false | 0.28 |
| false | true | true | 0.08 |
| true | false | false | 0.05 |
| true | false | true | 0.06 |
| true | true | false | 0.07 |
| true | true | true | 0.02 |

Full joint probability distribution

# PROBABILISTIC GRAPHICAL MODELS (PGMS)

- Semantics given by (normalised) product of $m$ local factors




Factor graph (FG)

$$P_F = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(R_{i,1}, \dots, R_{i,k_i})$$

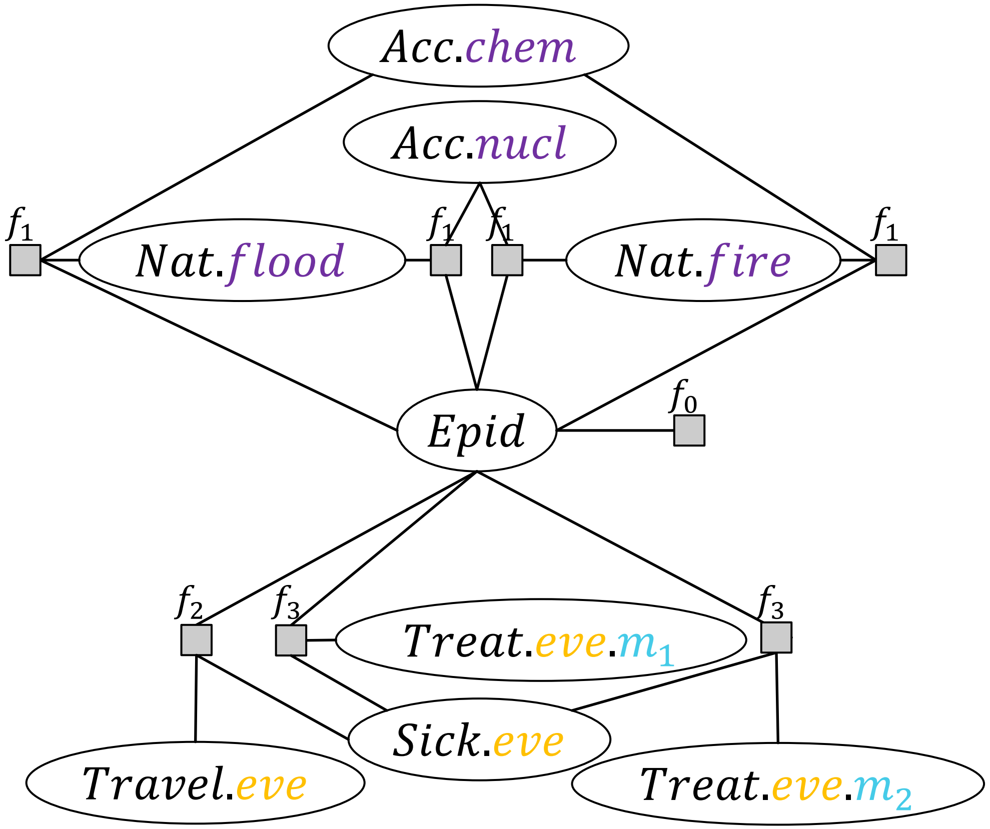
$$Z = \sum_{r_1 \in \mathrm{ran}(R_1)} \dots \sum_{r_n \in \mathrm{ran}(R_n)} \prod_{i=1}^{m} \phi_i(r_{i,1}, \dots, r_{i,k_i})$$

| $Epid$ | $Travel$ | $Sick$ | $P$ |
|---|---|---|---|
| false | false | false | 0.20 |
| false | false | true | 0.24 |
| false | true | false | 0.28 |
| false | true | true | 0.08 |
| true | false | false | 0.05 |
| true | false | true | 0.06 |
| true | true | false | 0.07 |
| true | true | true | 0.02 |

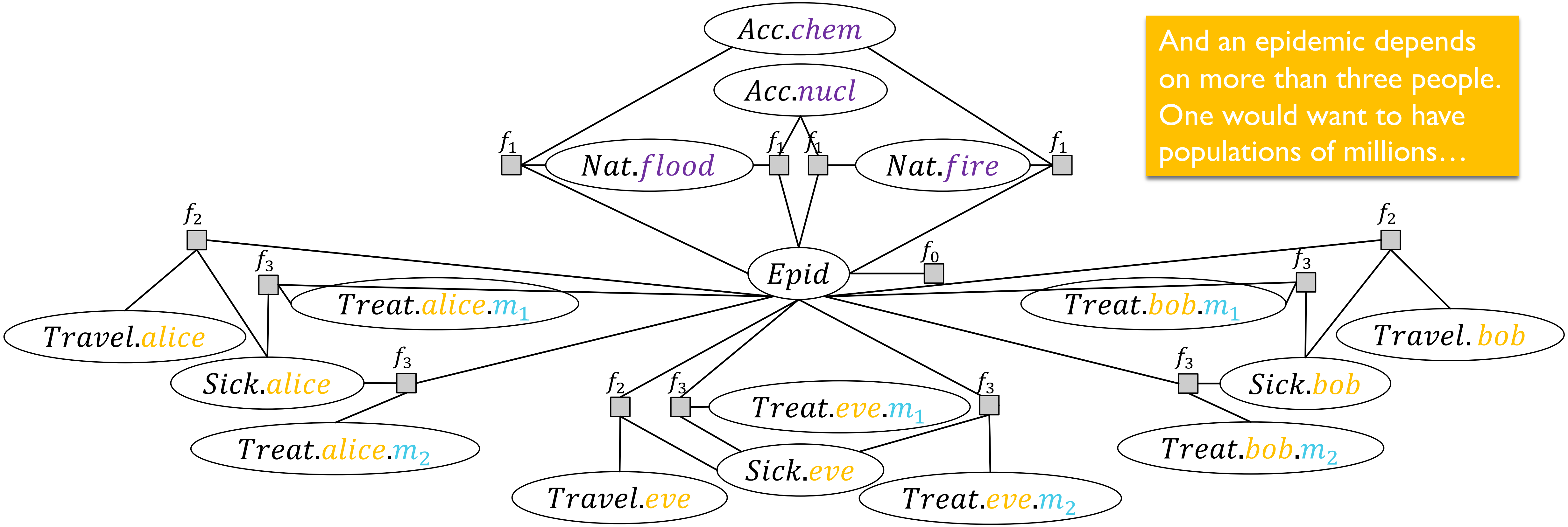
Full joint probability distribution

- Probabilistic reasoning: Exploit factorisation → push sums into the product
- Size exponential in the largest number of arguments $s$ of a factor: $O(m \cdot r^s)$
- Runtime exponential in the largest interim factor during reasoning: $O(m \cdot r^w)$

# PROBLEM: ADDING OBJECTS AND RELATIONS IN PROPOSITIONAL FORMALISMS CLUNKY…

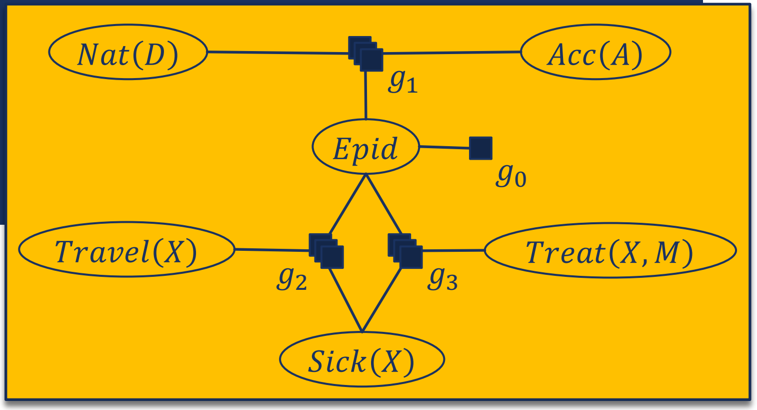


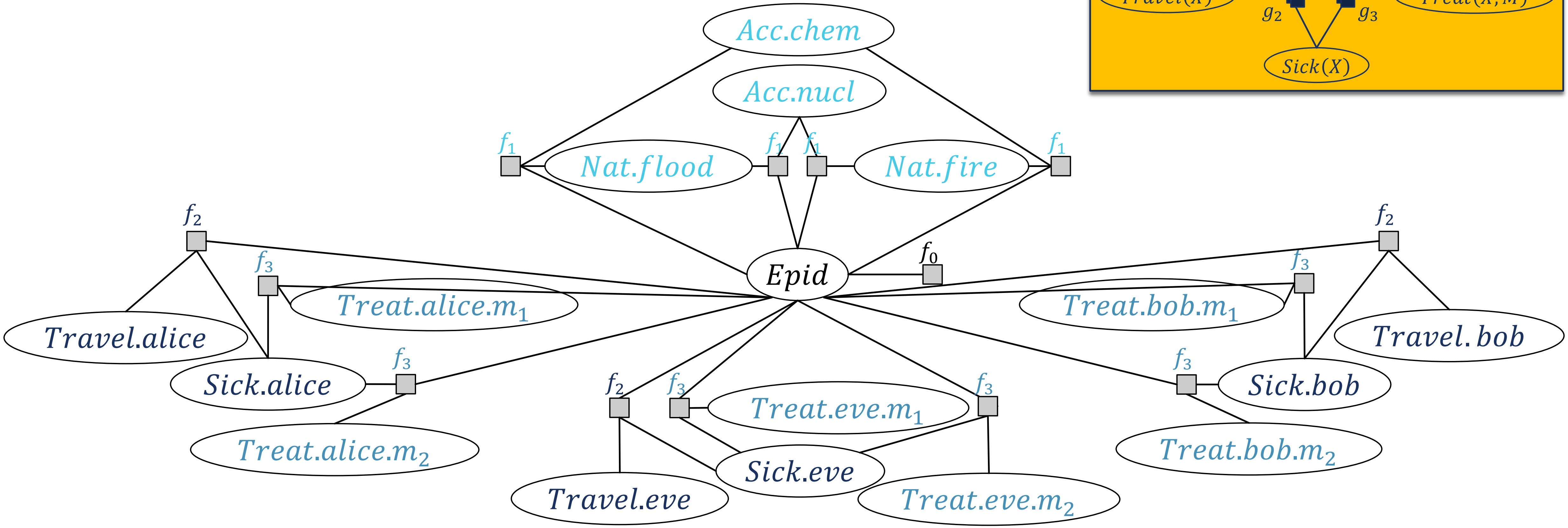
# PROBLEM: … AND MODELS EXPLODE WITH THEM




And an epidemic depends on more than three people. One would want to have populations of millions…

# PROPOSITIONAL → FIRST-ORDER VIEW




**Symmetries in graph / relations in scenario**

# DESCRIPTION OF THE ENVIRONMENT

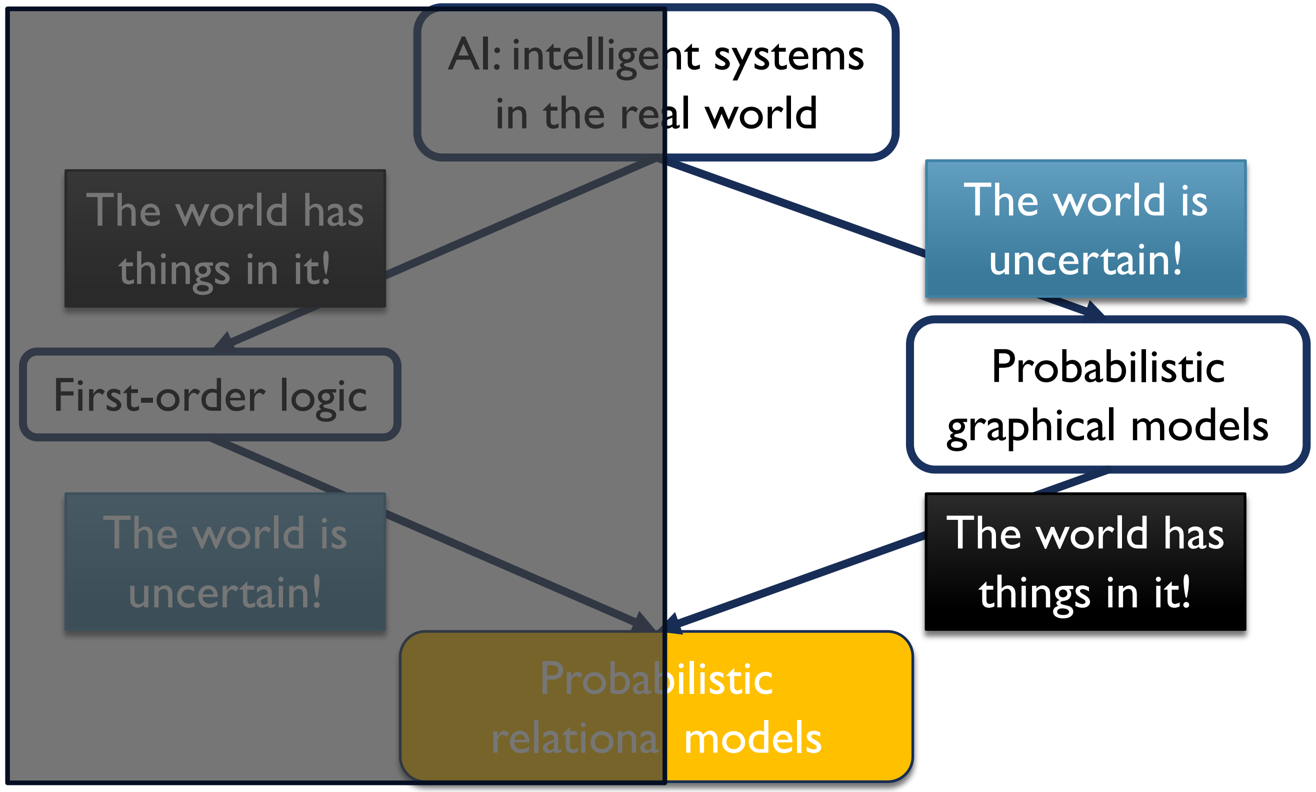



Figure based on a figure by Stuart Russell

# THE WORLD HAS THINGS IN IT: LOGIC

- Describe hard constraints about a world
- Components
  - Logical language

    *Andy likes Cody.* *Abby does **not** like Dana.* *Bess likes **everyone** that Bess likes.*
    *Dana does **not** like Abby.* *Bess likes Cody **or** Dana.*

  - Logical reasoning

| **Premises:** | **Truths:** | **Falsehoods:** | **Unknowns:** |
| --- | --- | --- | --- |
| *Dana likes Cody.* | *Bess likes Cody.* | *Bess likes Dana.* | *Dana likes Bess.* |
| *Abby does not like Dana.* | *Bess does not like Dana.* | *Everybody likes everybody.* | |
| *Everybody likes somebody.* | *Everybody likes someone.* | | |
| *Bess likes Cody or Dana.* | | | |
| *Abby likes everyone that Bess likes.* | | | |
| *Cody likes everyone who likes her.* | | | |
| *Nobody likes herself.* | | | |

- Set of constants (lowercase)
- Logical operators (¬,∧,∨, ⇒, ⇔)
- (Parentheses)

# PROPOSITIONAL → FIRST-ORDER LOGIC

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

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

- Motivation: Being able to talk about objects and relations among them
  - Example: propositional thinking
    - Premises:
      If Jack knows Jill, then Jill knows Jack. Jack knows Jill.
    - Conclusion: Does Jill know Jack? (Yes.)
  - Example: first-order (or relational) thinking
    - Premises:
      If one person knows another person, then the other person knows the first person. Jack knows Jill.
    - Conclusion: Does Jill know Jack? (Yes.)

- Example: in first-order (relational) logic
  - Premises
    - $\forall X. \forall Y. \left(knows(X, Y) \Rightarrow knows(Y, X)\right)$
    - $knows(jack, jill)$
  - Conclusion: $knows(jill, jack)$?

- (Logical) variables (uppercase)
- Quantifiers (∀, ∃)
- Set of constants (lowercase)
- Logical operators (¬,∧,∨, ⇒, ⇔)
- (Parentheses)

# DESCRIPTION OF THE ENVIRONMENT

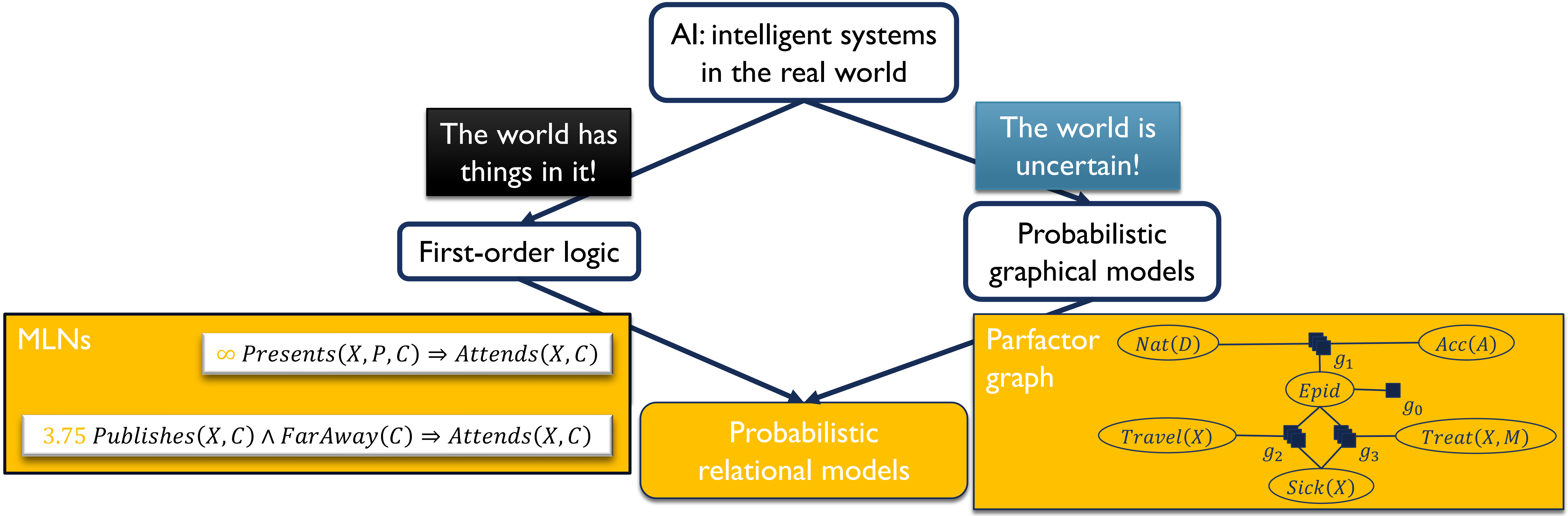



Figure based on a figure by Stuart Russell

# PROBABILISTIC (RELATIONAL) MODELS

PARFACTORS, MARKOV LOGIC NETWORKS

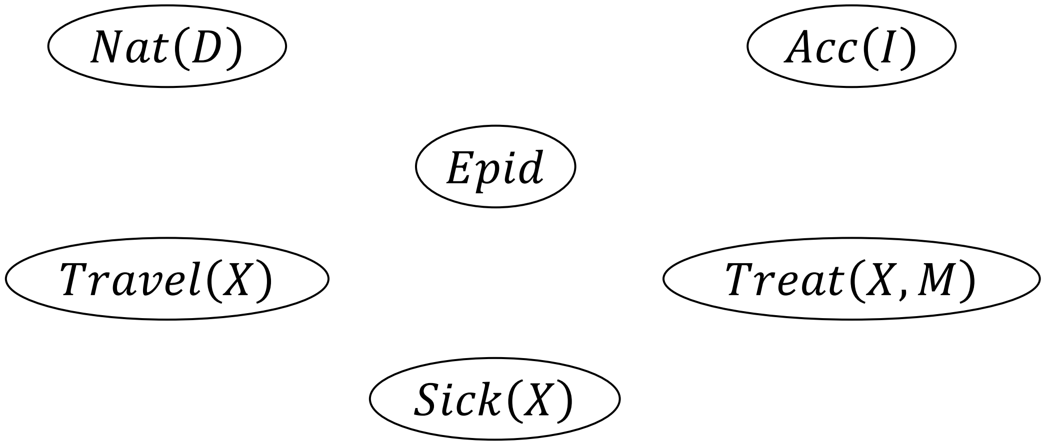
# LOGICAL VARIABLES IN RANDOM VARIABLES

- 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

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



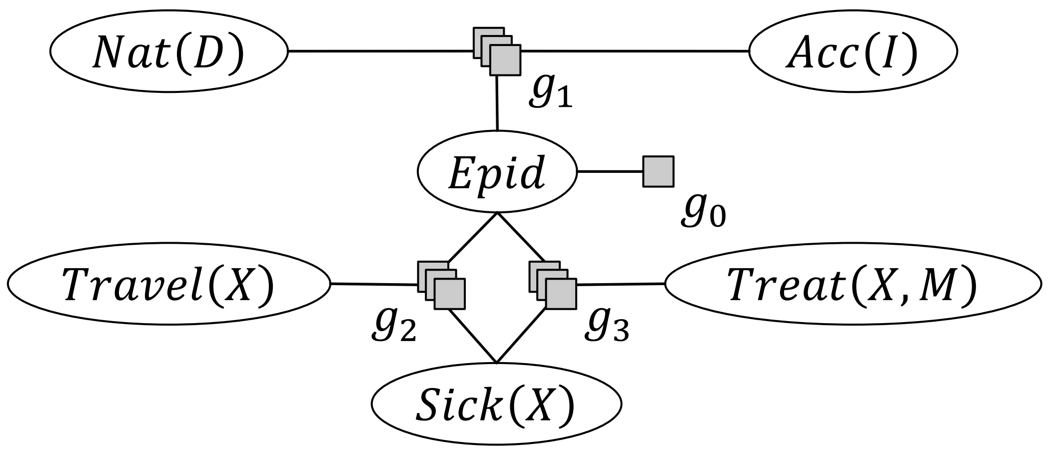
# PARFACTORS

- Factors with PRVs = parfactors

| $Travel(X)$ | $Epid$ | $Sick(X)$ | $g_2$ |
|---|---|---|---|
| $false$ | $false$ | $false$ | 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

| $Travel(eve)$ | $Epid$ | $Sick(eve)$ | $g_2$ |
|---|---|---|---|
| $false$ | $false$ | $false$ | 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 |

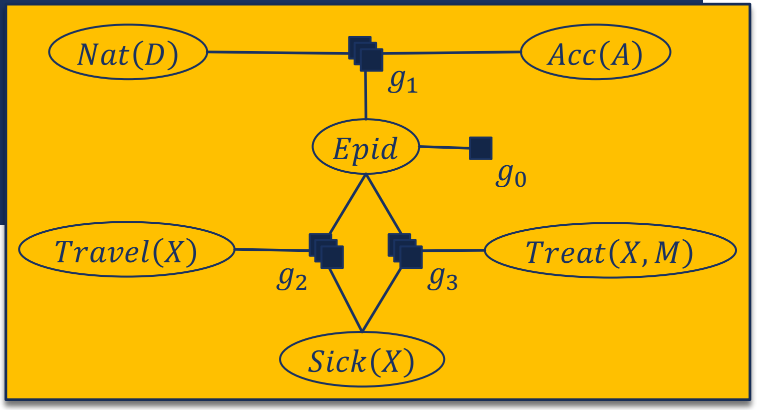
- Grounding

| $Travel(X)$ | $Epid$ | $Sick(X)$ | $g_2$ |
|---|---|---|---|
| $false$ | $false$ | $false$ | 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 |

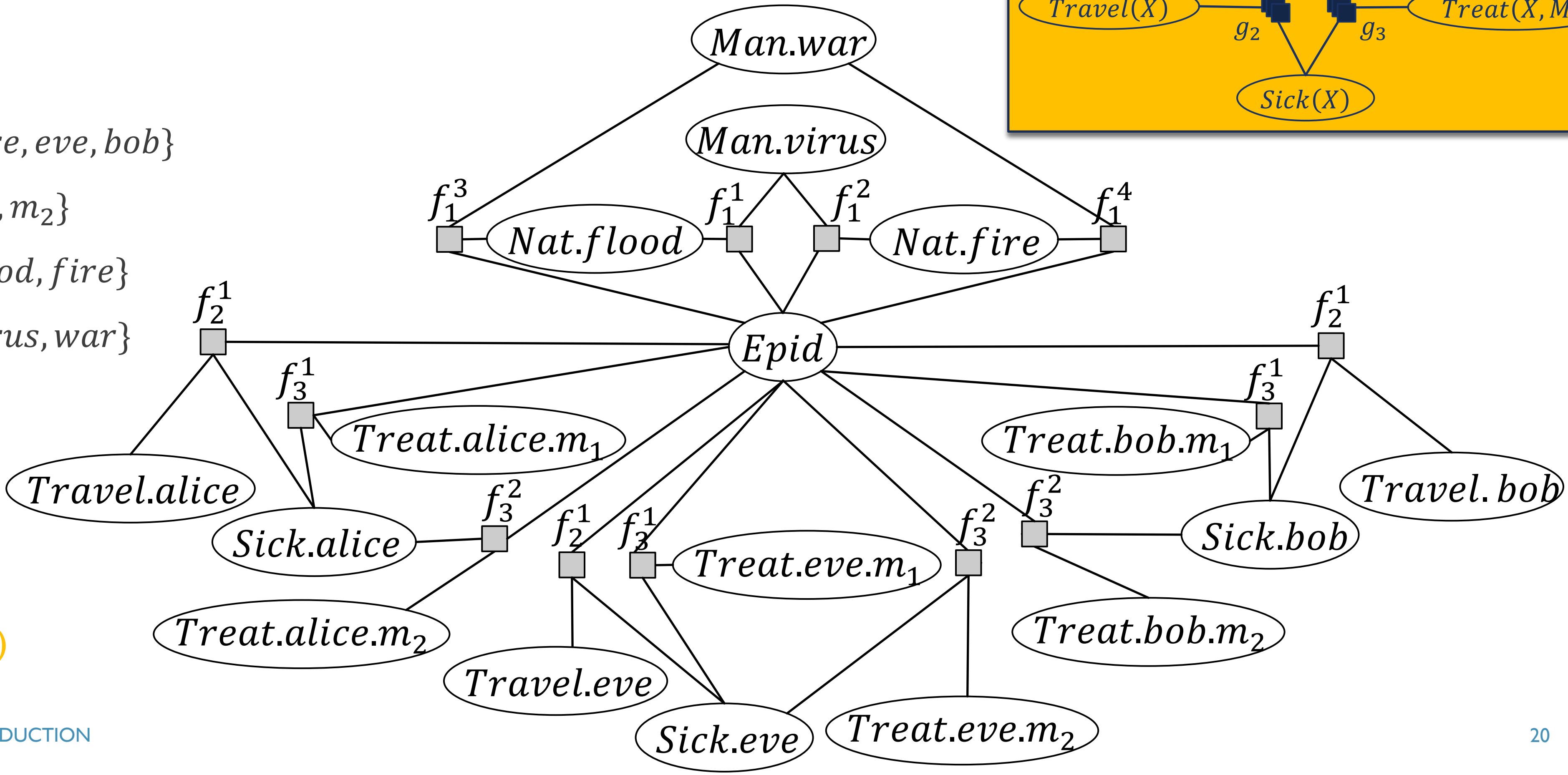
| $Travel(alice)$ | $Epid$ | $Sick(alice)$ | $g_2$ |
|---|---|---|---|
| $false$ | $false$ | $false$ | 5 |
| $false$ | $false$ | $true$ | 0 |
| $false$ | $true$ | $false$ | 4 |
| $false$ | $true$ | $true$ | 6 |
| $true$ | $false$ | $false$ | 4 |
| $true$ | $false$ | $true$ | 6 |
| $true$ | $true$ | $false$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

$at(X, M)$

# 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)

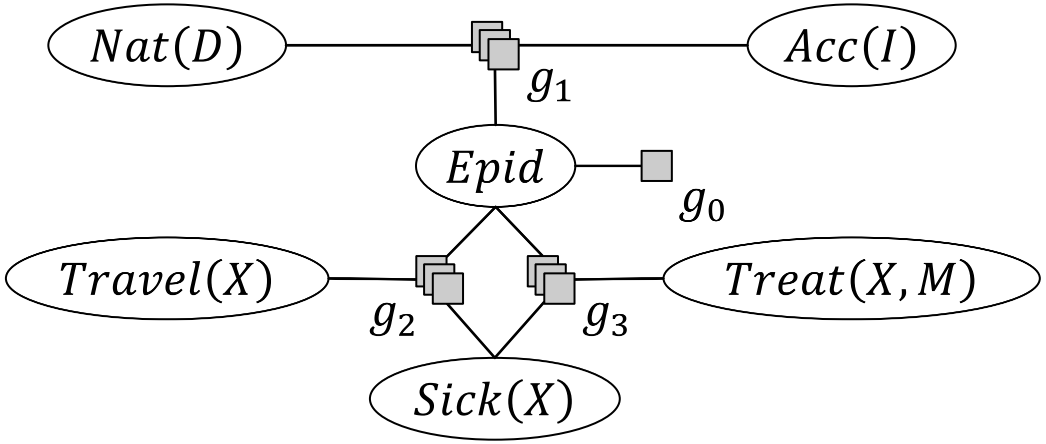


# SEMANTICS

- Given model $G = \{g_i\}_{i=1}^{n}$
  - Over PRVs $\mathrm{rvs}(G)$
  - Equivalent propositional model $F = \mathrm{gr}(G)$ over random variables $\mathrm{gr}\big(\mathrm{rv}(G)\big) = \{R_1, \ldots, R_N\} =: \boldsymbol{R}$ with semantics $P_{\boldsymbol{R}}$
- Semantics: Full joint probability distribution $P_G$

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

$$Z = \sum_{r_1 \in \mathrm{ran}(R_1)} \ldots \sum_{r_N \in \mathrm{ran}(R_N)} \prod_{f \in \mathrm{gr}(G)} f$$




  - Equivalence over semantics: $P_G = P_F = P_{\boldsymbol{R}}$

# SYMMETRIES WITHIN

- Assume four epidemics with identical characteristics
  - $Epid_1, Epid_2, Epid_3, Epid_4$
  - Reasonable to model the epidemics such that it does not matter which $Epid$ variables specifically are $true$ or $false$, i.e., they are interchangeable

    - All $false$ maps to $8$
    - 1 $true$, 3 $false$ maps to $6$
    - 2 $true$, 2 $false$ maps to $4$
    - 3 $true$, 1 $false$ maps to $2$
    - All $true$ maps to $0$

    → Five lines enough to describe

$\#_{true}$ $\#_{false}$

| | |
|---|---|
| $[0,4]$ | $8$ |
| $[1,3]$ | $6$ |
| $[2,2]$ | $4$ |
| $[3,1]$ | $2$ |
| $[4,0]$ | $0$ |

| $Epid_1$ | $Epid_2$ | $Epid_3$ | $Epid_4$ | $\phi$ |
|---|---|---|---|---|
| $false$ | $false$ | $false$ | $false$ | $8$ |
| $false$ | $false$ | $false$ | $true$ | $6$ |
| $false$ | $false$ | $true$ | $false$ | $6$ |
| $false$ | $false$ | $true$ | $true$ | $4$ |
| $false$ | $true$ | $false$ | $false$ | $6$ |
| $false$ | $true$ | $false$ | $true$ | $4$ |
| $false$ | $true$ | $true$ | $false$ | $4$ |
| $false$ | $true$ | $true$ | $true$ | $2$ |
| $true$ | $false$ | $false$ | $false$ | $6$ |
| $true$ | $false$ | $false$ | $true$ | $4$ |
| $true$ | $false$ | $true$ | $false$ | $4$ |
| $true$ | $false$ | $true$ | $true$ | $2$ |
| $true$ | $true$ | $false$ | $false$ | $4$ |
| $true$ | $true$ | $false$ | $true$ | $2$ |
| $true$ | $true$ | $true$ | $false$ | $2$ |
| $true$ | $true$ | $true$ | $true$ | $0$ |

# COUNTING RANDOM VARIABLE (CRV)

- New PRV type: (Parameterised) counting random variable ((P)CRV) $\#_X[A_{|C}]$
  - $A_{|C}$ a PRV under constraint $C$
  - $X \in \mathrm{lv}(A)$
  - Range values: Histogram $h = \{(v_i, n_i)\}_{i=1}^m$
    - $m = |\mathrm{ran}(A)|$ (number of buckets)
    - $n = \sum_{i=1}^m n_i = \left|\mathrm{gr}\left(A_{|\pi_X(C)}\right)\right|$ (number of instances to distribute into buckets)
    - $v_i \in \mathrm{ran}(A)$ (buckets)
    - $n_i \in \mathbb{N}$ (number of instances in bucket $v_i$)
    - Shorthand: $[n_1, \ldots, n_m]$
- Range of a (P)CRV = space of histograms fulfilling the conditions on the histograms
  - (All possible ways of distributing $n$ interchangeable instances into $m$ buckets)
- Single histogram encodes several interchangeable assignments at once
  - Given by multinomial coefficient $\mathrm{Mul}(h)$

$$\mathrm{Mul}(h) = \frac{n!}{\prod_{i=1}^m n_i!}$$

    - If $m = 2$, binomial coefficient:

$$\binom{n}{n_1} = \frac{n!}{(n-n_1)!\,n_1!} = \frac{n!}{n_2!\,n_1!}$$

# CRV: EXAMPLE

- (P)CRV $\#_X[A_{|C}]$
  - Range values: Histogram $h = \{(v_i, n_i)\}_{i=1}^m$
    - $m = |\mathrm{ran}(A)|$ (number of buckets)
    - $n = \sum_{i=1}^m n_i = |\mathrm{gr}(A_{|C})|$ (number of instances to distribute into buckets)
    - $v_i \in \mathrm{ran}(A)$ (buckets)
    - $n_i \in \mathbb{N}$ (number of instances in bucket $v_i$)
    - Shorthand: $[n_1, \dots, n_m]$
  - Single histogram encodes several interchangeable assignments at once: $\mathrm{Mul}(h) = \frac{n!}{\prod_{i=1}^m n_i!}$

- E.g., CRV: $\#_E[Epid(E)]$
  - $ran(Epid(E)) = \{true, false\} \rightarrow m = 2$
  - $dom(E) = \{e_1, e_2, e_3, e_4\} \rightarrow n = 4$
  - Range values and multiplicities

$$\{(true, 0), (false, 4)\} = [0,4] \quad Mul([0,4]) = \frac{4!}{0! \cdot 4!} = 1$$
$$\{(true, 1), (false, 3)\} = [1,3] \quad Mul([1,3]) = \frac{4!}{1! \cdot 3!} = 4$$
$$\{(true, 2), (false, 2)\} = [2,2] \quad Mul([2,2]) = \frac{4!}{2! \cdot 2!} = 6$$
$$\{(true, 3), (false, 1)\} = [3,1] \quad Mul([3,1]) = \frac{4!}{3! \cdot 1!} = 4$$
$$\{(true, 4), (false, 0)\} = [4,0] \quad Mul([4,0]) = \frac{4!}{4! \cdot 0!} = 1$$

# CRV: EXAMPLE

- E.g., (continued)
  - CRV: $\#_E[Epid(E)]$
    - Range values
      $[0,4]$, $[1,3]$, $[2,2]$, $[3,1]$, $[4,0]$
      1 4 6 4 1
      how many assignments encoded
  - $g' = \phi(\#_E[Epid(E)])$

| $\#_E[Epid(E)]$ | $\phi'$ |
|---|---|
| [0,4] | 8 |
| [1,3] | 6 |
| [2,2] | 4 |
| [3,1] | 2 |
| [4,0] | 0 |

| $Epid_1$ | $Epid_2$ | $Epid_3$ | $Epid_4$ | $\phi$ |
|---|---|---|---|---|
| $false$ | $false$ | $false$ | $false$ | 8 |
| $false$ | $false$ | $false$ | $true$ | 6 |
| $false$ | $false$ | $true$ | $false$ | 6 |
| $false$ | $false$ | $true$ | $true$ | 4 |
| $false$ | $true$ | $false$ | $false$ | 6 |
| $false$ | $true$ | $false$ | $true$ | 4 |
| $false$ | $true$ | $true$ | $false$ | 4 |
| $false$ | $true$ | $true$ | $true$ | 2 |
| $true$ | $false$ | $false$ | $false$ | 6 |
| $true$ | $false$ | $false$ | $true$ | 4 |
| $true$ | $false$ | $true$ | $false$ | 4 |
| $true$ | $false$ | $true$ | $true$ | 2 |
| $true$ | $true$ | $false$ | $false$ | 4 |
| $true$ | $true$ | $false$ | $true$ | 2 |
| $true$ | $true$ | $true$ | $false$ | 2 |
| $true$ | $true$ | $true$ | $true$ | 0 |

# CRVS CONTINUED

- (P)CRV $\#_X[A_{|C}]$ with
  - $m = |ran(A)|$ (number of buckets)
  - $n = \sum_{i=1}^{m} n_i = |gr(A_{|\pi_X(C)})|$ (number of instances to distribute into buckets)
- Instead of $m^n$ mappings in the ground factor, the counted factor has

$$\binom{n+m-1}{n-1}$$

mappings

  - Upper bound of range size of a CRV:

$$\binom{n+m-1}{n-1} \leq n^m$$

Exponential in number of random variables $n$

$\downarrow$

Polynomial in number of random variables $n$

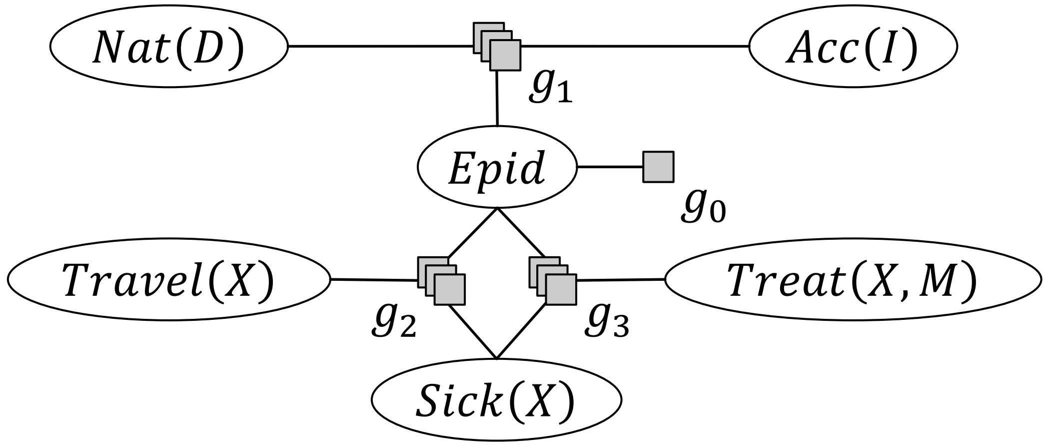
# SEMANTICS – REPRISE

Semantics stay the same
→ CRVs need to be grounded by *expanding the histograms into the mappings they encode*

- Given model $G = \{g_i\}_{i=1}^{n}$
  - Over PRVs $\mathrm{rvs}(G)$
  - Equivalent propositional model $F = \mathrm{gr}(G)$ over random variables $\mathrm{gr}\big(\mathrm{rv}(G)\big) = \{R_1, \dots, R_N\} =: \boldsymbol{R}$ with semantics $P_{\boldsymbol{R}}$
- Semantics: Full joint probability distribution $P_G$

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

$$Z = \sum_{r_1 \in \mathrm{ran}(R_1)} \dots \sum_{r_N \in \mathrm{ran}(R_N)} \prod_{f \in \mathrm{gr}(G)} f$$




  - Equivalence over semantics: $P_G = P_F = P_{\boldsymbol{R}}$

# MARKOV LOGIC NETWORKS (MLNs)

- Again: First-order view on probabilistic modelling
- Use logical formulas to specify potential functions
  - Weights for each formula induce a full joint again
    - Lower weight if constraint does not hold all the time
    - ∞ weight for hard constraints, in practice either a high weight or only worlds considered that satisfy hard constraints

Hard constraint

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

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

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

# GROUNDING

- Each $(w_i, \psi_i)$ represents a set of *propositional* sentences, each sentence with weight $w_i$
  - One sentence for each possible substitution of the free variables $\text{free}(\psi_i)$ in $\psi_i$ given a finite domain (or a constraint set) $D$ over $\text{free}(\psi_i)$
    - $\theta_D = \bigcup_{\boldsymbol{d} \in D} \{\bigcup_{t \in \boldsymbol{d}} \{X_d \to t\}\}$
  - Example: MLN $\Psi = \{(w_i, \psi_i)\}_{i=1}^{2}$
    - Domains
      - $\text{dom}(X) = \{\text{alice}\}$
      - $\text{dom}(P) = \{\text{p}_1, \text{p}_2\}$
      - $\text{dom}(C) = \{\text{ijcai}, \text{ki}\}$
    - Groundings on the right

$$\big(10, \mathit{Presents}(\text{alice}, \text{p}_1, \text{ijcai}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ijcai})\big)$$
$$\big(10, \mathit{Presents}(\text{alice}, \text{p}_1, \text{ki}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ki})\big)$$
$$\big(10, \mathit{Presents}(\text{alice}, \text{p}_2, \text{ijcai}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ijcai})\big)$$
$$\big(10, \mathit{Presents}(\text{alice}, \text{p}_2, \text{ki}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ki})\big)$$
$$\big(3.75, \mathit{Publishes}(\text{alice}, \text{ijcai}) \wedge \mathit{FarAway}(\text{ijcai}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ijcai})\big)$$
$$\big(3.75, \mathit{Publishes}(\text{alice}, \text{ki}) \wedge \mathit{FarAway}(\text{ki}) \Rightarrow \mathit{Attends}(\text{alice}, \text{ki})\big)$$

$$\boxed{\infty\ \mathit{Presents}(X, P, C) \Rightarrow \mathit{Attends}(X, C)}$$

$$\boxed{3.75\ \mathit{Publishes}(X, C) \wedge \mathit{FarAway}(C) \Rightarrow \mathit{Attends}(X, C)}$$

# MLNS: SEMANTICS

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

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

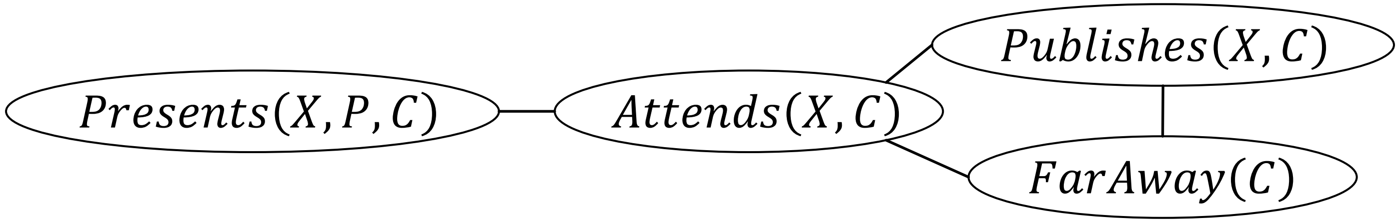
  - $N$ = the number of ground atoms in the grounded $\Psi$
  - Probability of one interpretation $\omega$

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

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

# 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, let us build an analogue:
  - Logical atoms as nodes
  - Edges between atoms whenever atoms occur together in a formula
    - Form cliques in graph
    - Potential functions (factors) for cliques from weights
  - Can also build an equivalent parfactor graph as graphical representation


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

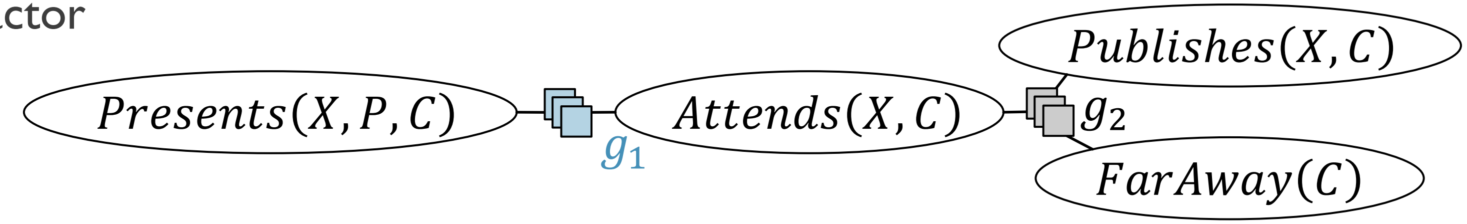
$$3.75 \; Publishes(X,C) \land FarAway(C) \Rightarrow Attends(X,C)$$

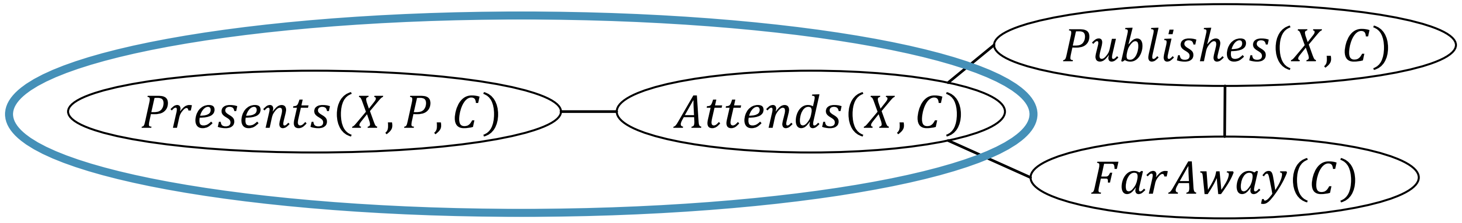
# FROM WEIGHTED FORMULAS TO PARFACTORS

An MLN can be translated into a set of parfactors and vice versa [Van den Broeck 13]

- For each weighted formula, build a parfactor
  - Atoms as arguments
  - For assignment $\boldsymbol{a}$ that makes $\psi_i$ true, map $\boldsymbol{a}$ to $\exp(w_i)$

| $Pr(X,P,C)$ | $Att(X,C)$ | $\phi_1$ |
|---|---|---|
| false | false | exp 10 |
| false | true | exp 10 |
| true | false | exp 0 |
| true | true | exp 10 |




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

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

# A SELECTION OF OTHER PROBABILISTIC RELATIONAL MODELS

## Distribution semantics

- Probabilistic extensions to Datalog [Fuhr 1995]
- Relational Bayesian networks [Jaeger 1997]
- Bayesian Logic Programming [Milch et al. 2005], ProbLog [De Raedt et al. 2007]
- Parfactor models [Poole 03, Taghipour et al. 2013]
- MLNs [Richardson & Domingos 2006]

## Other semantics

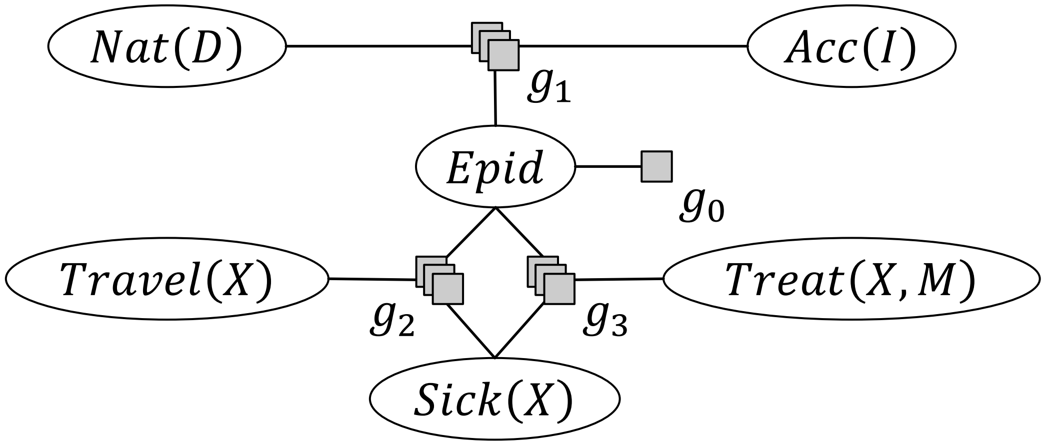
- Hinge-loss Markov networks, Probabilistic Soft Logic (PSL) [Bach et al. 2017]
  - Define density function using log-linear model
- Maximum entropy semantics [Thimm et al. 2010]
  - Partial specification of discrete joint with "uniform completion"

# PROBABILISTIC REASONING IN PROBABILISTIC RELATIONAL MODELS

## EXPLOITING SYMMETRIES FOR EFFICIENCY

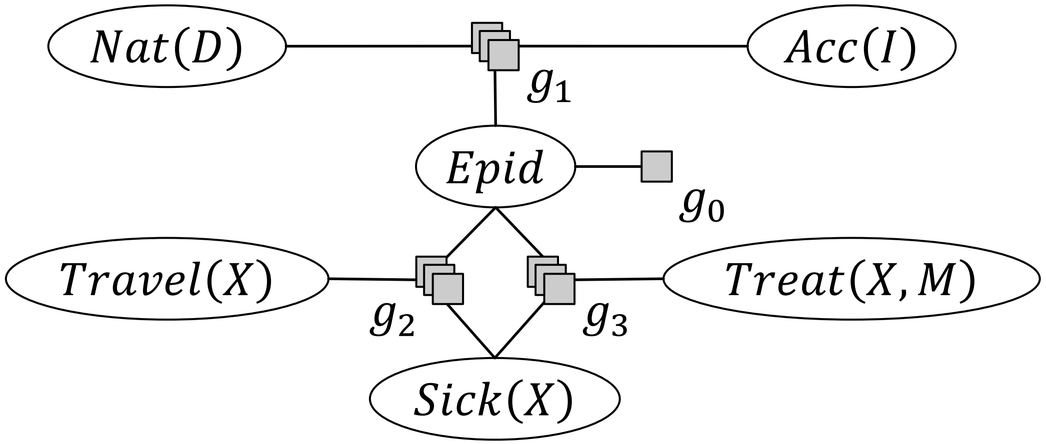
# INFERENCE PROBLEMS WITH AND WITHOUT EVIDENCE

- Query answering problem (QA) given a model:
  - Probability of events, e.g., $P(Att(eve, ki) = true), P(Epid = true)$
  - Conditional (marginal) probability distributions, e.g., $P(Att(ev, ki)|FarAway(ki)), P(Epid|sick(alice), sick(eve))$
  - 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



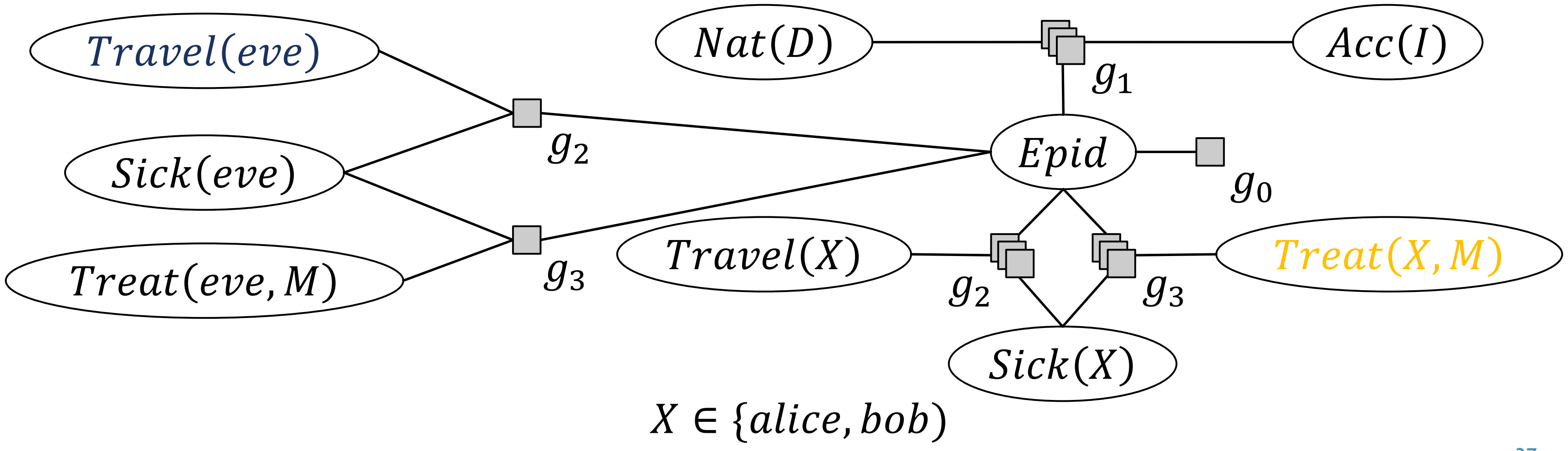
# QA IN PARFACTOR MODELS: LIFTED VARIABLE ELIMINATION (LVE)

- Eliminate all variables not appearing in query
  - [Poole 2003, de Salvo Braz et al. 2005+2006, Milch et al. 2008, Taghipour et al. 2013, B & Möller 2018]
- 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

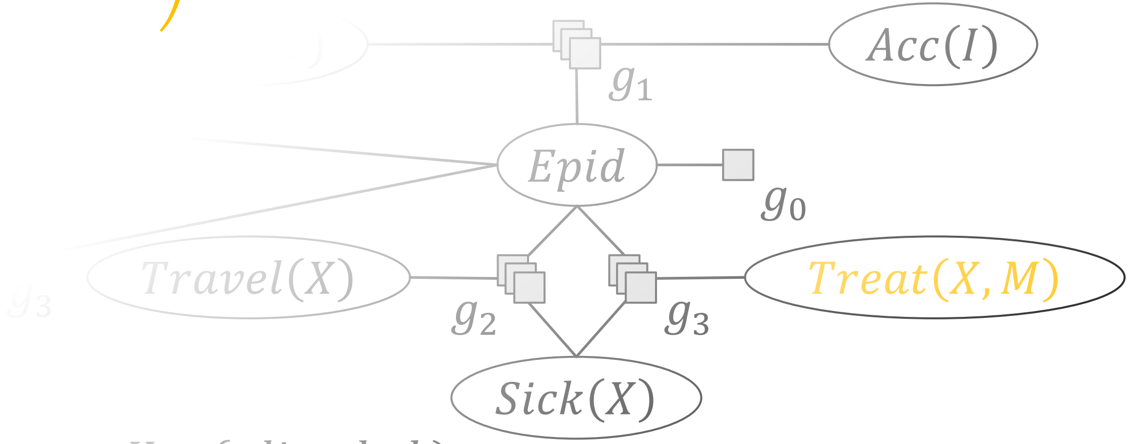
- Eliminate $Treat(X, M)$
  - Appears in only one $g$: $g_3$
  - Contains all logical variables of $g_3$: $X, M$
  - For each $X$ constant: the same number of $M$ constants
  - ✓ Preconditions of lifted summing out fulfilled, lifted summing out possible



# LVE IN DETAIL: LIFTED SUMMING OUT

- Eliminate $Treat(X, M)$ by lifted summing out
  1. Sum out representative

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




$X \in \{alice, bob\}$

# LVE IN DETAIL: LIFTED SUMMING OUT

| $Epid$ | $Sick(X)$ | $Treat(X,M)$ | $g_3$ |
|---|---|---|---|
| $false$ | $false$ | $false$ | 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 |

| $Epid$ | $Sick(X)$ | $\Sigma$ |
|---|---|---|
| $false$ | $false$ | 10 |
| $false$ | $true$ | 9 |
| $true$ | $false$ | 12 |
| $true$ | $true$ | 12 |

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

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

# A SELECTION OF OTHER LIFTED INFERENCE ALGORITHMS

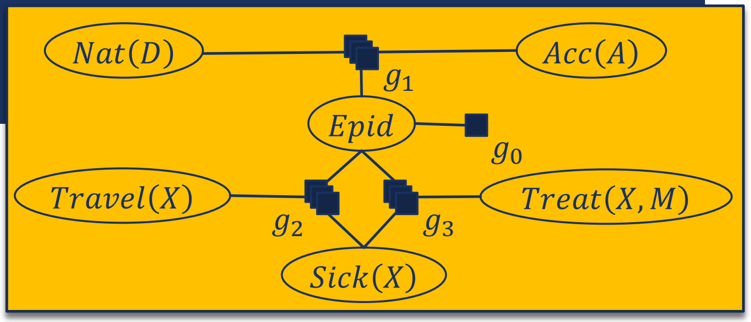
- For multiple queries
  - Parfactor graphs: Lifted junction tree algorithm [B and Möller 2016]
  - MLNs: First-order knowledge compilation [Van den Broeck et al. 2011], probabilistic theorem proving [Gogate & Domingos 2011]
- Approximate inference
  - Symmetric factor graphs: Lifted belief propagation [Ahmadi et al. 2013]
  - MLNs: Lifted importance sampling [Venegupal & Gogate 2014]
  - Symmetries within Markov chains: Lifted MCMC [Niepert 2012]
  - Factor graphs: Local symmetries [Holtzen et al. 2019]

# COMPLEXITY & TRACTABILITY

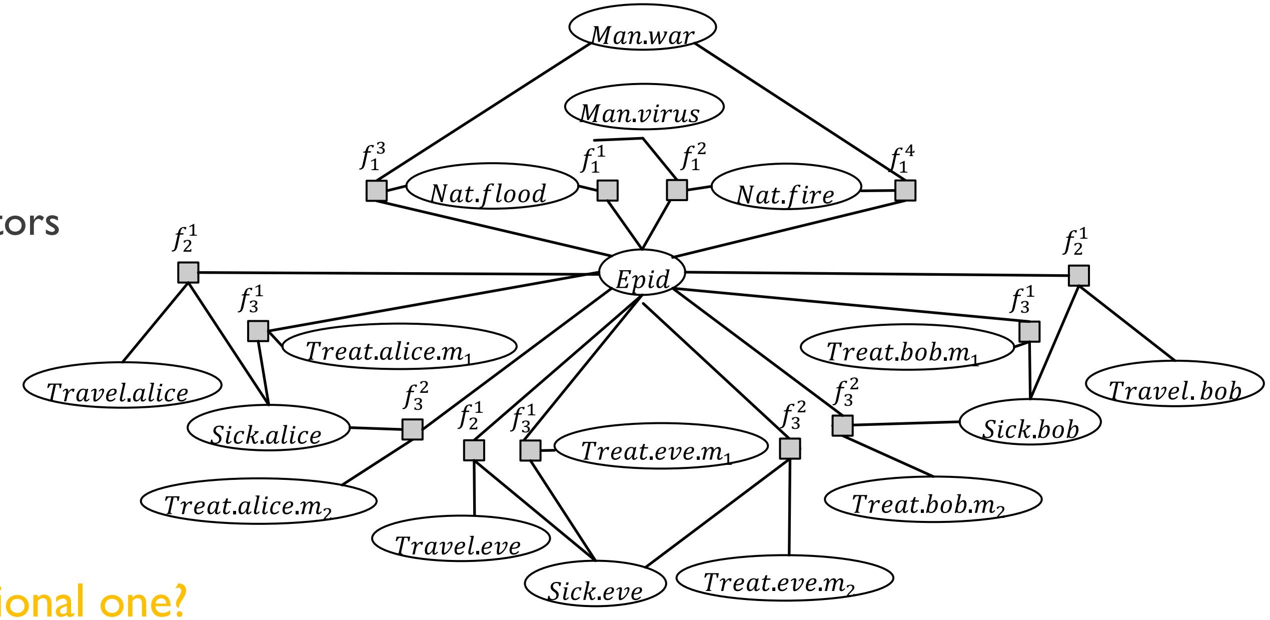
- Given a model that allows for lifted calculations
  - I.e., no groundings during solving an instance of the problem
- Solving an instance of the problem is possible in time polynomial in domain sizes

  → The query answering algorithm is domain-lifted
- A query answering problem is tractable
  - when it is solved by an efficient algorithm, running in time polynomial in the number of random variables
- Assume that the number of random variables is 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]

# SUMMARY




- (Exchangeable) objects and relations in an uncertain world
- Probabilistic relational models
  - Parfactor graphs: Adding logical variables to factors
  - MLNs: Adding weights to logical formulas
  - Grounding semantics with symmetries in weights / factors
- Lifted inference
  - Exploit symmetries for efficient inference
  - Tractability by exchangeability
- **Open problem:**
  How do we get a first-order model from a propositional one?



# AGENDA

1. Introduction and background [Tanya]
2. Compressing models [Malte]
   - The basics of colour passing
   - Recent advances in colour passing
3. Accuracy considerations [Jan]
4. Summary and future directions [Tanya]

# BIBLIOGRAPHY

## ORDERED ALPHABETICALLY

# BIBLIOGRAPHY


- Ahmadi et al. [2013]
  Babak Ahmadi, Kristian Kersting, Martin Mladenov, and Sriraam Natarajan: Exploiting Symmetries for Scaling Loopy Belief Propagation and Relational Training. In *Machine Learning*, 2013.
- AMAI, Russel/Norvig
  Russell & Norvig: Artificial Intelligence: A Modern Approach. 2020.
- 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 [2016]
  Tanya Braun and Ralf Möller: Lifted Junction Tree Algorithm. In *KI-16 Proceedings of the 39th German Conference on Artificial Intelligence*, 2016.

# BIBLIOGRAPHY


- 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.
- 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.
- 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.

# BIBLIOGRAPHY


- 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.
- 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.
- Holtzen et al. [2019]
  Steven Holtzen, Todd Millstein, and Guy Van den Broeck: Generating and Sampling Orbits for Lifted Probabilistic Inference. In *UAI-19 Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, 2019
- Jaeger [1997]
  Manfred Jaeger: Relational Bayesian Networks. In *UAI-97 Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, 1997.

# BIBLIOGRAPHY


- 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.
- Niepert [2012]
  Mathias Niepert: Markov Chains on Orbits of Permutation Groups. In *UAI-2012 Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.
- 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.

# BIBLIOGRAPHY


- Poole [2003]
  David Poole: First-order Probabilistic Inference. In *IJCAI-03 Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003.
- Richardson & Dominogs [2006]
  Matthew Richardson and Pedro Domingos: Markov Logic Networks. In *Machine Learning*, 2006.
- Taghipour et al. [2013]
  Nima Taghipour, Daan Fierens, Jesse Davis, and Hendrik Blockeel. Lifted Variable Elimination: Decoupling the Operators from the Constraint Language. In *Journal of Artificial Intelligence Research*, 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.

# BIBLIOGRAPHY


- 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.
- Venegupal & Gogate [2014]
  Deepak Venugopal and Vibhav Gogate: Scaling-up Importance Sampling for Markov Logic Networks. In *Advances in Neural Information Processing Systems 27*, 2014.