

Fundamental Problems in Statistical Relational AI

Tutorial - KR 2024

TU Wien, Austria
Machine Learning Research Unit

StarAI

ooooooooo

Tractability

oooooooooooooooooooooooooooooooooooo

Consistency

oooooooooooooooooooo

TABLE OF CONTENTS

Statistical Relational AI

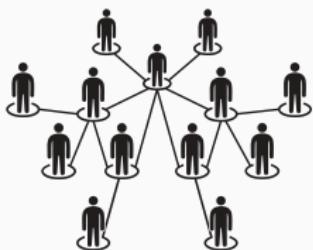
Tractability

Consistency

StarAI

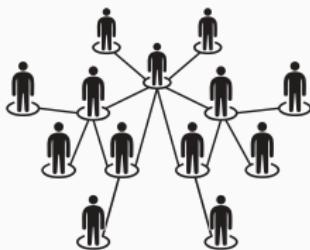
WHAT ARE RELATIONAL DOMAINS?

Social Networks

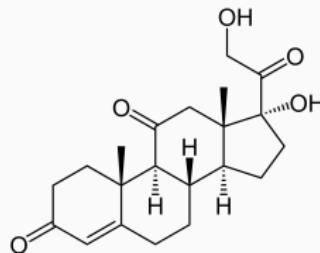


WHAT ARE RELATIONAL DOMAINS?

Social Networks

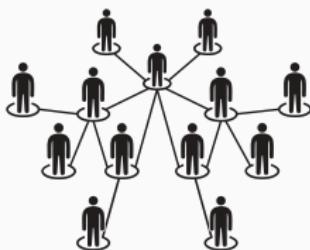


Molecules

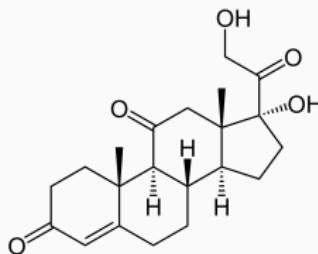


WHAT ARE RELATIONAL DOMAINS?

Social Networks



Molecules

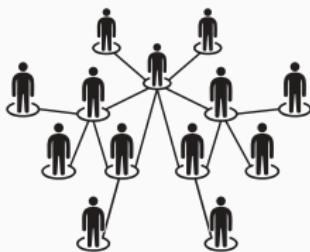


Business Processes



MODELING RELATIONAL DOMAINS IN FIRST ORDER LOGIC

Social Networks



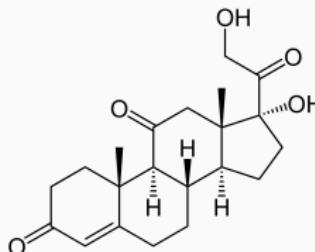
Friends(x, y)

Smokes(x)

Cancer(x)

•

Molecules



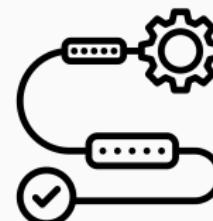
Bond(x, y)

Carbon(x)

Oxygen(x)

20

Business Processes



Teaches(x, y)

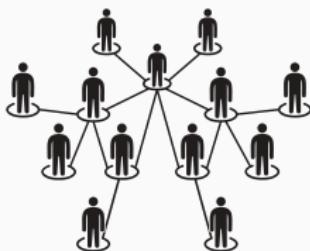
Professor(x)

Student(x)

1

MODELING KNOWLEDGE IN FIRST ORDER LOGIC

Social Networks

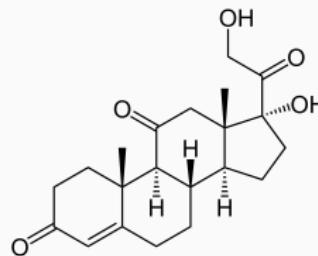


$$\forall xy. \text{Fr}(x,y) \rightarrow \text{Fr}(y,x)$$

$$\forall xy. \text{Sm}(x) \wedge \text{Fr}(x,y) \rightarrow \text{Sm}(y)$$

•
•
•

Molecules



$$\forall x. \text{H}(x) \rightarrow \exists y. \text{Bond}(x, y)$$

$\forall x. 0(x) \rightarrow \exists^{\leq 2} y. \text{Bond}(x, y)$

•

Business Processes



$$\forall xy. \text{Prof}(x) \wedge \text{Tch}(x,y) \rightarrow \text{Stud}(y)$$

MODELLING RELATIONAL DOMAINS IN FIRST ORDER LOGIC

PROBLEMS!¹

- **Laziness:** Too much work to list out all rules!

¹Russell and Norvig. Artificial Intelligence: A Modern Approach

MODELLING RELATIONAL DOMAINS IN FIRST ORDER LOGIC

PROBLEMS!¹

- **Laziness**: Too much work to list out all rules!
- **Theoretical Ignorance**: We don't have all the rules!

¹Russell and Norvig. Artificial Intelligence: A Modern Approach

MODELLING RELATIONAL DOMAINS IN FIRST ORDER LOGIC

PROBLEMS!¹

- **Laziness**: Too much work to list out all rules!
- **Theoretical Ignorance**: We don't have all the rules!
- **Practical Ignorance**: Maybe there are no rules — inherent stochasticity!

¹Russell and Norvig. Artificial Intelligence: A Modern Approach

STATISTICAL RELATIONAL AI

Statistical Relational AI

二

Logic + Probability

INGREDIENTS OF STATISTICAL RELATIONAL LEARNING

SRL ingredients:

- A set of Herbrand models $\Omega^{(n)}$ in function-free First Order Logic
- A parametric probability distribution $\mathbb{P}_\theta : \Omega \rightarrow [0, 1]$

INGREDIENTS OF STATISTICAL RELATIONAL LEARNING

SRL ingredients:

- A set of Herbrand models $\Omega^{(n)}$ in function-free First Order Logic
- A parametric probability distribution $\mathbb{P}_\theta : \Omega \rightarrow [0, 1]$

The Inference Problem:

- $\mathbb{P}_\theta(q) = \sum_{q \models \omega} \mathbb{P}_\theta(\omega)$

INGREDIENTS OF STATISTICAL RELATIONAL LEARNING

SRL ingredients:

- A set of Herbrand models $\Omega^{(n)}$ in function-free First Order Logic
- A parametric probability distribution $\mathbb{P}_\theta : \Omega \rightarrow [0, 1]$

The Inference Problem:

- $\mathbb{P}_\theta(q) = \sum_{q \models \omega} \mathbb{P}_\theta(\omega)$

The Learning Problem:

- $\theta^* = \operatorname{argmax}_\theta \mathbb{P}_\theta(\omega^*)$

FUNDAMENTAL PROBLEMS IN SRL

“What are the fundamental problems in Statistical Relational Learning?”

FUNDAMENTAL PROBLEMS IN SRL

“What are the fundamental problems in Statistical Relational Learning?”

The Inference Problem:

- $\mathbb{P}_\theta(q) = \sum_{q \models \omega} \mathbb{P}_\theta(\omega)$ **Intractability**

FUNDAMENTAL PROBLEMS IN SRL

“What are the fundamental problems in Statistical Relational Learning?”

The Inference Problem:

$$\cdot \mathbb{P}_\theta(q) = \sum_{q \models \omega} \mathbb{P}_\theta(\omega)$$

Intractability

The Learning Problem:

$$\cdot \theta^* = \operatorname{argmax}_\theta \mathbb{P}_\theta(\omega^*)$$

No Consistent Estimation

Tractability

MARKOV LOGIC²: AN EXAMPLE

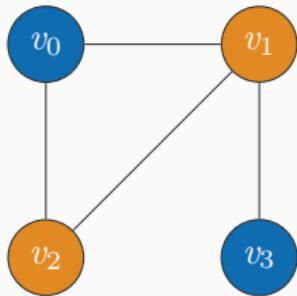
A contact-tracing model:

$$w_1 : \text{Covid}(x)$$

$$w_2 : \text{Covid}(x) \wedge \text{Contact}(x, y) \rightarrow \text{Covid}(y)$$

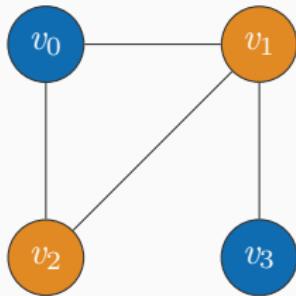
²Richardson and Domingos. Markov Logic Networks. 2006

MLNs: AN EXAMPLE



- **Covid(x)**
 $n_1(\omega) = |\{v_i : \omega \models \text{Covid(v}_i\text{)}\}|$

MLNs – AN EXAMPLE



- $\text{Covid}(x) \wedge \text{Contact}(x, y) \rightarrow \text{Covid}(y)$
- $n_2(\omega) = |\{(v_i, v_j) : \omega \models \phi_2(v_i, v_j)\}|$

MARKOV LOGIC NETWORKS: ALMOST FORMALLY

- Weighted quantifier-free first-order logic formulas:

$$\{w_i : \phi_i\}$$

MARKOV LOGIC NETWORKS: ALMOST FORMALLY

- Weighted quantifier-free first-order logic formulas:

$$\{w_i : \phi_i\}$$

- Probability distribution on all finite structures of size n :

$$\Pr(\omega) := \frac{1}{Z} \exp \left(\sum_i w_i n_i(\omega) \right) \quad (1)$$

MARKOV LOGIC NETWORKS: ALMOST FORMALLY

- Weighted quantifier-free first-order logic formulas:

$$\{w_i : \phi_i\}$$

- Probability distribution on all finite structures of size n :

$$\Pr(\omega) := \frac{1}{Z} \exp \left(\sum_i w_i n_i(\omega) \right) \quad (1)$$

- The partition function – **main source of intractability**:

$$Z := \sum_{\omega \models \Phi_\infty} \exp \left(\sum_i w_i n_i(\omega) \right) \quad (2)$$

WEIGHTED FIRST ORDER MODEL COUNTING

Symmetric Weighted First Order Model Counting (WFOMC):

$$\text{WFOMC}(\Phi, n) := \sum_{\omega \models \Phi} w(\omega)$$

WEIGHTED FIRST ORDER MODEL COUNTING

Symmetric Weighted First Order Model Counting (WFOMC):

$$\text{WFOMC}(\Phi, n) := \sum_{\omega \models \Phi} w(\omega)$$

$$w(\omega) = \prod_{\substack{\omega \models g \\ g \in \mathcal{G}}} w(pred(g)) \prod_{\substack{\omega \models \neg g \\ g \in \mathcal{G}}} \bar{w}(pred(g)).$$

WEIGHTED FIRST ORDER MODEL COUNTING

Symmetric Weighted First Order Model Counting (WFOMC):

$$\text{WFOMC}(\Phi, n) := \sum_{\omega \models \Phi} w(\omega)$$

$$w(\omega) = \prod_{\substack{\omega \models g \\ g \in \mathcal{G}}} w(pred(g)) \prod_{\substack{\omega \models \neg g \\ g \in \mathcal{G}}} \bar{w}(pred(g)).$$

WFOMC is **Tractable** if it can be computed in polynomial time w.r.t the domain size (n).

WFOMC: AN EXAMPLE

$$\forall x. KR(x) \rightarrow Intelligent(x)$$

$$w(KR) = 2.5$$

$$\bar{w}(KR) = 0.5$$

$$w(Intelligent) = 0.5$$

$$\bar{w}(Intelligent) = 1.5$$

KR(a)	Intelligent(a)	w(ω)
1	1	$2.5 \times 0.5 = 1.25$
1	0	$2.5 \times 1.5 = 1.25$
0	1	$5 \times 0.5 = 1.5$
0	0	$5 \times 1.5 = 7.5$

$$WFOMC(\Phi, 1) = 1.25 + 1.5 + 7.5$$

WFOMC \equiv PARTITION FUNCTION

$$Z := \sum_{\omega \models \Phi_\infty} \exp \left(\sum_i w_i n_i(\omega) \right)$$

$$\Phi_\infty \wedge \bigwedge_i \forall FV[\phi_i]. (R_i(FV[\phi_i]) \leftrightarrow \phi_i) \quad (3)$$

$$w(R_i) = \exp(w_i)$$

$$\bar{w}(R_i) = 1$$

$$w(*) = 1$$

$$\bar{w}(*) = 1$$

WFOMC \equiv PARTITION FUNCTION: AN EXAMPLE

$$w : \text{KR}(x) \rightarrow \text{Intelligent}(x)$$

WFOMC \equiv PARTITION FUNCTION: AN EXAMPLE

$$w : \text{KR}(x) \rightarrow \text{Intelligent}(x)$$

WFOMC encoding for the partition function:

$$\forall x. R(x) \leftrightarrow (\text{KR}(x) \rightarrow \text{Intelligent}(x))$$

$$w(R) = \exp(w)$$

$$w(*) = 1$$

$$\bar{w}(*) = 1$$

HOW INTRACTABLE (PRACTICALLY)?

WFOMC is a #P complete problem in general³.

³Beame, Van den Broeck, Gribkoff, Suciu. PODS 2015.

HOW INTRACTABLE (PRACTICALLY)?

WFOMC is a #P complete problem in general³.

But is WMC not practically quite scalable?

³Beame, Van den Broeck, Gribkoff, Suciu. PODS 2015.

How INTRACTABLE?

 $w_1 : \text{Stress}(x) \rightarrow \text{Smokes}(x)$ $w_2 : \text{Smokes}(x) \wedge \text{Fr}(x, y) \rightarrow \text{Smokes}(y)$ $w_3 : \text{Fr}(x, y) \wedge \text{Fr}(y, z) \rightarrow \text{Fr}(x, z)$

HOW INTRACTABLE⁴?

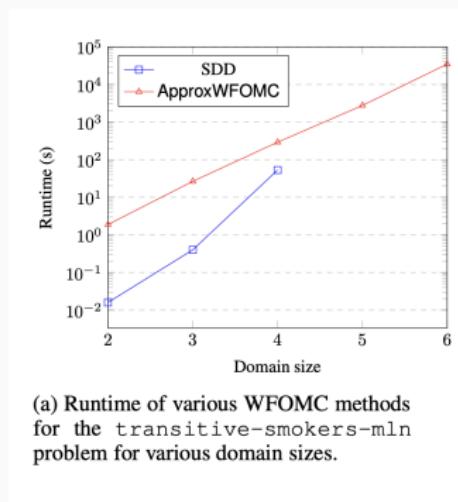


Figure 1: 10000 seconds is more than 2 hours

⁴Bremen and Kuzelka. IJCAI 2020

WFOMC: FUNDAMENTAL PROBLEMS

What fragments of first-order logic admit tractable WFOMC?

⁵Beame, Van den Broeck, Gribkoff and Suciu. PODS 2015.

⁶den Broeck et al. KR 2014

WFOMC: FUNDAMENTAL PROBLEMS

What fragments of first-order logic admit tractable WFOMC?

There is an intractable FO^3 formula⁵

⁵Beame, Van den Broeck, Gribkoff and Suciu. PODS 2015.

⁶den Broeck et al. KR 2014

WFOMC: FUNDAMENTAL PROBLEMS

What fragments of first-order logic admit tractable WFOMC?

There is an intractable FO^3 formula⁵

FO^2 is tractable⁶

⁵Beame, Van den Broeck, Gribkoff and Suciu. PODS 2015.

⁶den Broeck et al. KR 2014

FO² LANGUAGE: 1-TYPES

We have a language with **at most two variables**, with the following predicates:

- A unary predicate **KR(x)**
- A binary predicate **Shaves(x, y)**

FO² LANGUAGE: 1-TYPES

We have a language with **at most two variables**, with the following predicates:

- A unary predicate $KR(x)$
- A binary predicate $Shaves(x, y)$

We have the following set of unary properties also called **1-types**:

$$\neg KR(c) \wedge \neg Shaves(c, c)$$

$$\neg KR(c) \wedge Shaves(c, c)$$

$$KR(c) \wedge \neg Shaves(c, c)$$

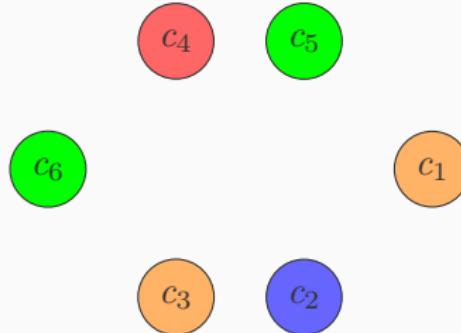
$$KR(c) \wedge Shaves(c, c)$$

1- TYPE ENUMERATION

An arrangement of 1-Types

$$\neg \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$$

$$\neg \text{KR}(c) \wedge \text{Shaves}(c, c)$$



$$k : \quad k_1 = 1 \quad k_2 = 2 \quad k_3 = 1 \quad k_4 = 2$$

$$\text{KR}(c) \wedge \neg \text{Shaves}(c, c)$$

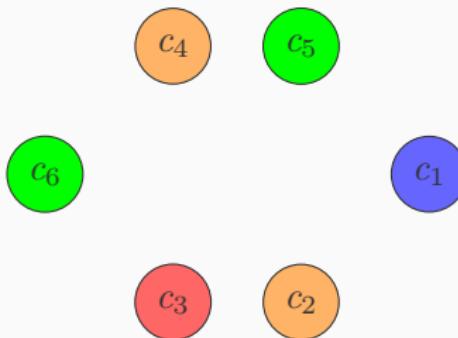
$$\text{KR}(c) \wedge \text{Shaves}(c, c)$$

1- TYPE ENUMERATION

Another arrangement of 1-Types

$$\neg \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$$

$$\neg \text{KR}(c) \wedge \text{Shaves}(c, c)$$



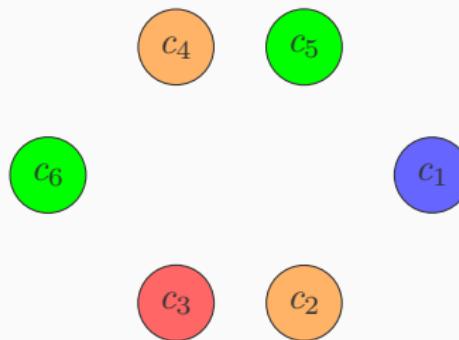
$$k : \quad k_1 = 1 \quad k_2 = 2 \quad k_3 = 1 \quad k_4 = 2$$

$$\text{KR}(c) \wedge \neg \text{Shaves}(c, c)$$

$$\text{KR}(c) \wedge \text{Shaves}(c, c)$$

1-TYPE ENUMERATION

Counting for fixed 1-Type cardinalities



$$k : \quad k_1 = 1 \quad k_2 = 2 \quad k_3 = 1 \quad k_4 = 2$$

$$\#\text{Similar Arrangements} = \binom{n}{k_1, k_2, k_3, k_4} = \frac{6!}{1!2!1!2!} = 180$$

FO² LANGUAGE: 2-TABLES

We have an FO² language, with the following predicates:

- A unary predicate KR(x)
- A binary predicate Shaves(x, y)

We have the following set of binary properties also called **2-tables**:

Shaves(c, d) \wedge Shaves(d, c)

\neg Shaves(c, d) \wedge Shaves(d, c)

Shaves(c, d) \wedge \neg Shaves(d, c)

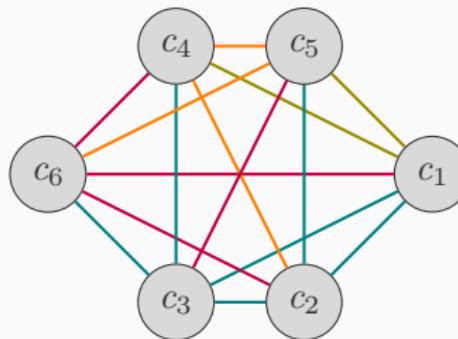
\neg Shaves(c, d) \wedge \neg Shaves(d, c)

2- TABLE ENUMERATION

An arrangement for 2-tables

$\text{Shaves}(c, d) \wedge \text{Shaves}(d, c)$

$\neg\text{Shaves}(c, d) \wedge \text{Shaves}(d, c)$

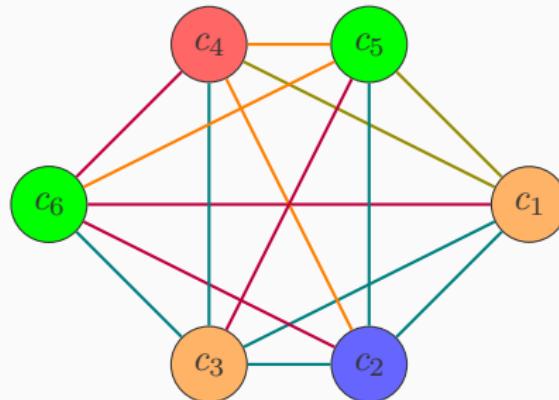


$\text{Shaves}(c, d) \wedge \neg\text{Shaves}(d, c)$

$\neg\text{Shaves}(c, d) \wedge \neg\text{Shaves}(d, c)$

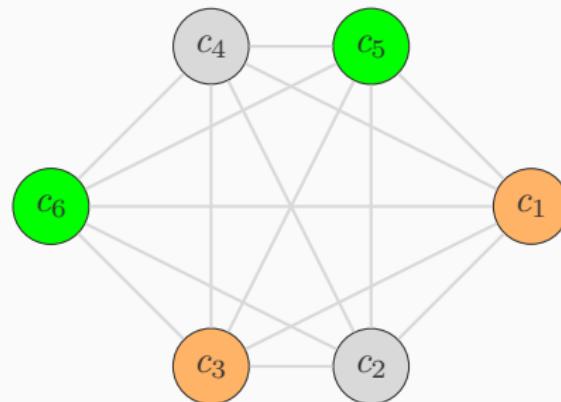
2- TABLE ENUMERATION GIVEN 1-TYPES

An arrangement for 2-tables given 1-types



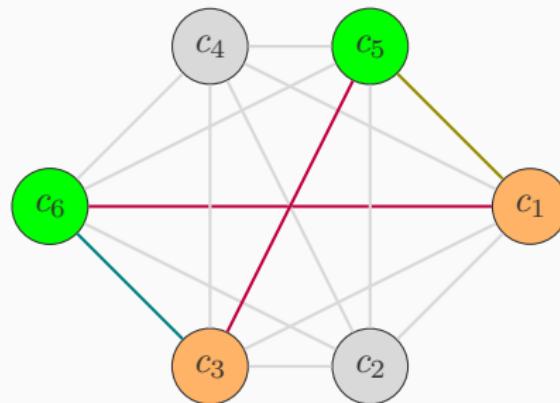
2- TABLE ENUMERATION GIVEN 1-TYPES

Pick a pair of 1-Types



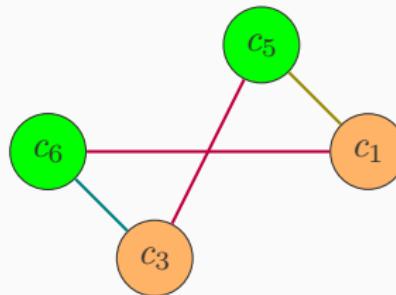
2- TABLE ENUMERATION GIVEN 1-TYPES

Picking a sub graph: Pick 2-Tables between them



2- TABLE ENUMERATION

Enumerating 2-tables given 1-types

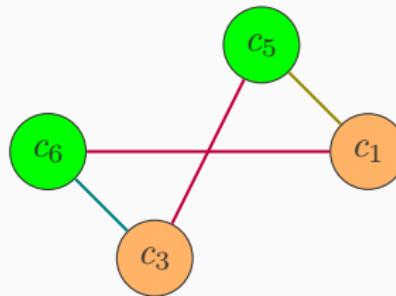


$$k_2 = 2 \quad k_4 = 2$$

$$h_1 = 2 \quad h_2 = 1 \quad h_3 = 1 \quad h_4 = 0$$

2- TABLE ENUMERATION

Enumerating 2-tables given 1-types



$$k_2 = 2 \quad k_4 = 2$$

$$h_1 = 2 \quad h_2 = 1 \quad h_3 = 1 \quad h_4 = 0$$

$$\binom{k_2 \times k_4}{h_1 \quad h_2 \quad h_3 \quad h_4} = \frac{(2 \times 2)!}{2!2!1!0!} = 6$$

ENUMERATING ALL MODELS OVER 1-TYPES AND 2-TABLES

$$\sum_{\langle \vec{k}, \vec{h} \rangle} \binom{n}{k_1 \dots k_u} \prod_{1 \leq i \leq j \leq u} \binom{k(i,j)}{h_1^{ij} \dots h_b^{ij}}$$

$$k(i,j) = \begin{cases} k_i k_j & \text{if } i \neq j \\ \frac{k_i(k_i-1)}{2} & \text{if } i = j \end{cases}$$

ADDING FORMULAS : $\forall xy.\Phi(x, y)$

A formula $\forall xy.\Phi(x, y)$, allows some and disallows other 1-Type and 2-Type configuration. For Example:

$$\neg \text{Shaves}(x, x)$$

$$\text{Shaves}(x, y) \rightarrow \text{Shaves}(y, x)$$

$$\text{KR}(x) \wedge \text{Shaves}(x, y) \rightarrow \text{KR}(y)$$

ADDING FORMULAS : $\forall xy.\Phi(x, y)$

$\neg \text{Shaves}(x, x)$

$\text{Shaves}(x, y) \rightarrow \text{Shaves}(y, x)$

$\text{KR}(x) \wedge \text{Shaves}(x, y) \rightarrow \text{KR}(y)$

Not allowed:

$\neg \text{KR}(c) \wedge \neg \text{Sh}(c, c) \wedge \text{KR}(d) \wedge \neg \text{Sh}(d, d) \wedge \text{Sh}(c, d) \wedge \text{Sh}(d, c)$



Allowed:

$\text{KR}(c) \wedge \neg \text{Sh}(c, c) \wedge \text{KR}(d) \wedge \neg \text{Sh}(d, d) \wedge \text{Sh}(c, d) \wedge \text{Sh}(d, c)$



FOMC IN FO^2 : $\forall x \forall y. \Phi(x, y)$

$$\text{FOMC}(\Phi, n) =$$

$$\sum_{\langle \vec{k}, \vec{h} \rangle} \binom{n}{k_1, \dots, k_u} \prod_{1 \leq i \leq j \leq u} \binom{\mathbf{k}(i,j)}{h_1^{ij}, \dots, h_b^{ij}} \prod_{1 \leq v \leq b} n_{ijv} h_v^{ij}$$

Unary Properties

Constraints: Φ

Binary Properties

FOMC IN FO^2 : $\forall x \forall y. \Phi(x, y)$

$$\text{FOMC}(\Phi, n) =$$

$$\begin{aligned}
 & \sum_{\langle \vec{\mathbf{k}}, \vec{\mathbf{h}} \rangle} \binom{n}{k_1, \dots, k_u} \prod_{1 \leq i \leq j \leq u} \binom{\mathbf{k}(i, j)}{h_1^{ij}, \dots, h_b^{ij}} \prod_{1 \leq v \leq b} n_{ijv}^{h_v^{ij}} \\
 &= \sum_{\langle \vec{\mathbf{k}}, \vec{\mathbf{h}} \rangle} \text{FOMC}(\Phi, \langle \vec{\mathbf{k}}, \vec{\mathbf{h}} \rangle) \\
 &= \sum_{\vec{\mathbf{k}}} \text{FOMC}(\Phi, \vec{\mathbf{k}})
 \end{aligned}$$

CARDINALITY CONSTRAINTS

$\neg \text{Shaves}(x, x)$

$\text{Shaves}(x, y) \rightarrow \text{Shaves}(y, x)$

$\text{KR}(x) \wedge \text{Shaves}(x, y) \rightarrow \text{KR}(y)$

$|\text{KR}(x)| = 2$

CARDINALITY CONSTRAINTS

$\neg \text{Shaves}(x, x)$

$\text{Shaves}(x, y) \rightarrow \text{Shaves}(y, x)$

$\text{KR}(x) \wedge \text{Shaves}(x, y) \rightarrow \text{KR}(y)$

$|\text{KR}(x)| = 2$

$k_1 : \neg \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$

$k_2 : \neg \text{KR}(c) \wedge \text{Shaves}(c, c)$

$k_3 : \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$

$k_4 : \text{KR}(c) \wedge \text{Shaves}(c, c)$

CARDINALITY CONSTRAINTS

$\neg \text{Shaves}(x, x)$

$\text{Shaves}(x, y) \rightarrow \text{Shaves}(y, x)$

$\text{KR}(x) \wedge \text{Shaves}(x, y) \rightarrow \text{KR}(y)$

$|\text{KR}(x)| = 2$

$k_1 : \neg \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$

$k_2 : \neg \text{KR}(c) \wedge \text{Shaves}(c, c)$

$k_3 : \text{KR}(c) \wedge \neg \text{Shaves}(c, c)$

$k_4 : \text{KR}(c) \wedge \text{Shaves}(c, c)$

$$k : k_3 + k_4 = 2$$

CARDINALITY CONSTRAINTS

Given an arbitrary cardinality constraint ρ :

$$\text{FOMC}(\Phi \wedge \rho) := \sum_{\langle k, h \rangle \models \rho} \text{FOMC}(\Phi, \langle k, h \rangle)$$

EXISTENTIAL QUANTIFIERS⁷

Minimal Non-Trivial Example:

$$\Psi := \Phi \wedge \forall x \exists y. R(x, y)$$

⁷Guy Van den Broeck, Wannes Meert, Adnan Darwiche. KR 2014
Malhotra and Serrafini. AAAI 2022 (for this formulation)

EXISTENTIAL QUANTIFIERS⁷

Minimal Non-Trivial Example:

$$\Psi := \Phi \wedge \forall x \exists y. R(x, y)$$

Key Proof ideas:

- $A_c = \{\omega | \omega \models \Phi \wedge \forall y. \neg R(c, y)\} \equiv$ Atleast c is isolated

⁷Guy Van den Broeck, Wannes Meert, Adnan Darwiche. KR 2014
Malhotra and Serrafini. AAAI 2022 (for this formulation)

EXISTENTIAL QUANTIFIERS⁷

Minimal Non-Trivial Example:

$$\Psi := \Phi \wedge \forall x \exists y. R(x, y)$$

Key Proof ideas:

- $A_c = \{\omega | \omega \models \Phi \wedge \forall y. \neg R(c, y)\} \equiv \text{Atleast } c \text{ is isolated}$
- WFOMC(Φ) - $|\bigcup_{i \in [n]} A_i|$

⁷Guy Van den Broeck, Wannes Meert, Adnan Darwiche. KR 2014
Malhotra and Serrafini. AAAI 2022 (for this formulation)

EXISTENTIAL QUANTIFIERS ⁷

Minimal Non-Trivial Example:

$$\Psi := \Phi \wedge \forall x \exists y. R(x, y)$$

Key Proof ideas:

- $A_c = \{\omega | \omega \models \Phi \wedge \forall y. \neg R(c, y)\} \equiv \text{Atleast } c \text{ is isolated}$
- WFOMC(Φ) - $|\bigcup_{i \in [n]} A_i|$
- Use Principle of Inclusion Exclusion to calculate $|\bigcup_{i \in [n]} A_i|$

⁷Guy Van den Broeck, Wannes Meert, Adnan Darwiche. KR 2014
Malhotra and Serrafini. AAAI 2022 (for this formulation)

PRINCIPLE OF INCLUSION-EXCLUSION

$$\cdot \left| \bigcup_{i=1}^n A_i \right| = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left| \bigcap_{j \in J} A_j \right|$$

PRINCIPLE OF INCLUSION-EXCLUSION

- $|\bigcup_{i=1}^n A_i| = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left| \bigcap_{j \in J} A_j \right|$
- The number of terms involved is $2^n - 1$
- Recall, $A_c = \{\omega \mid \omega \models \Phi \wedge \forall y. \neg R(c, y)\}$

PRINCIPLE OF INCLUSION-EXCLUSION

- $|\bigcup_{i=1}^n A_i| = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left| \bigcap_{j \in J} A_j \right|$
- The number of terms involved is $2^n - 1$
- Recall, $A_c = \{\omega \mid \omega \models \Phi \wedge \forall y. \neg R(c, y)\}$
- $\bigcap_{j \in J} A_j$ is dependent on $|J|$ and not on $\{J\}$

PRINCIPLE OF INCLUSION-EXCLUSION

- $|\bigcup_{i=1}^n A_i| = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} |\bigcap_{j \in J} A_j|$
- The number of terms involved is $2^n - 1$
- Recall, $A_c = \{\omega \mid \omega \models \Phi \wedge \forall y. \neg R(c, y)\}$
- $\bigcap_{j \in J} A_j$ is dependent on $|J|$ and not on $\{J\}$
- Let $\alpha_K = |\bigcap_{j \in J} A_j|$
- $\alpha_k = |\{\omega \mid \omega \models \Phi \wedge \forall xy. P(x) \rightarrow \neg R(x, y) \wedge |P| = k\}|$

PRINCIPLE OF INCLUSION-EXCLUSION

Recall,

$$\alpha_k = |\{\omega | \omega \models \Phi \wedge \forall xy. P(x) \rightarrow \neg R(x, y) \wedge |P| = k\}|$$

PRINCIPLE OF INCLUSION-EXCLUSION

Recall,

$$\alpha_k = |\{\omega | \omega \models \Phi \wedge \forall xy.P(x) \rightarrow \neg R(x, y) \wedge |P| = k\}|$$

We wanted:

$$\text{WFOMC}(\Phi) - |\bigcup_{i \in [n]} A_i|$$

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{k=1}^n (-1)^{k+1} \alpha_k$$

$$\left| U \setminus \bigcup_{i=1}^n A_i \right| = \sum_{k=0}^n (-1)^{k+1} \alpha_k$$

EXISTENTIAL QUANTIFIER \equiv PIE \equiv WFOMC WITH NEGATIVE WEIGHTS

$$\text{WFOMC}(\Phi \wedge \forall x \exists y. R(x, y)) \equiv \text{WFOMC}(\Phi \wedge \forall xy. P(x) \rightarrow \neg R(x, y))$$

EXISTENTIAL QUANTIFIER \equiv PIE \equiv WFOMC WITH NEGATIVE WEIGHTS

$$\text{WFOMC}(\Phi \wedge \forall x \exists y. R(x, y)) \equiv \text{WFOMC}(\Phi \wedge \forall xy. P(x) \rightarrow \neg R(x, y))$$

Scott's Normal Form is equi-satisfiable and WFOMC preserving:

$$\Phi \bigwedge_i \forall x \exists y R_i(x, y)$$

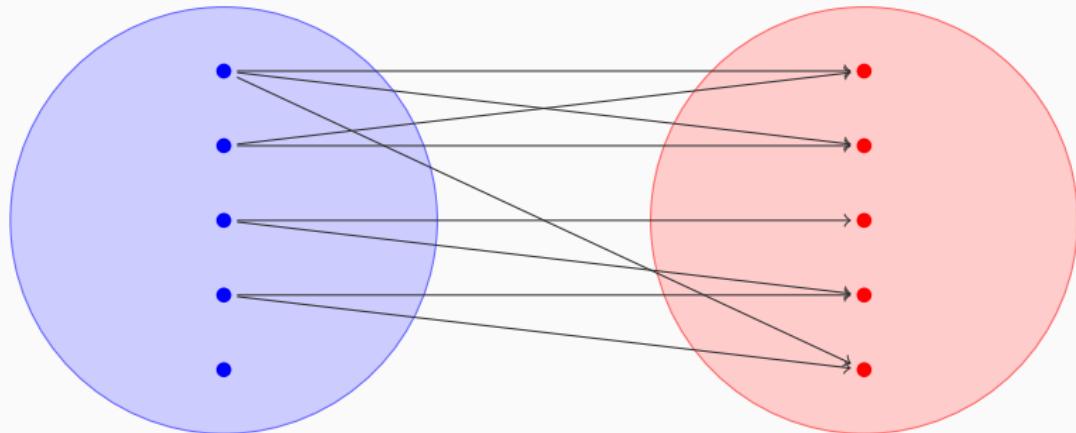
COUNTING QUANTIFIERS⁸

Counting Quantifiers: $\forall x \exists^{=1} y. R(x, y)$

⁸Kuzelka. JAIR 2021. Functionality: Kuuisto, Lutz. LICS 2018

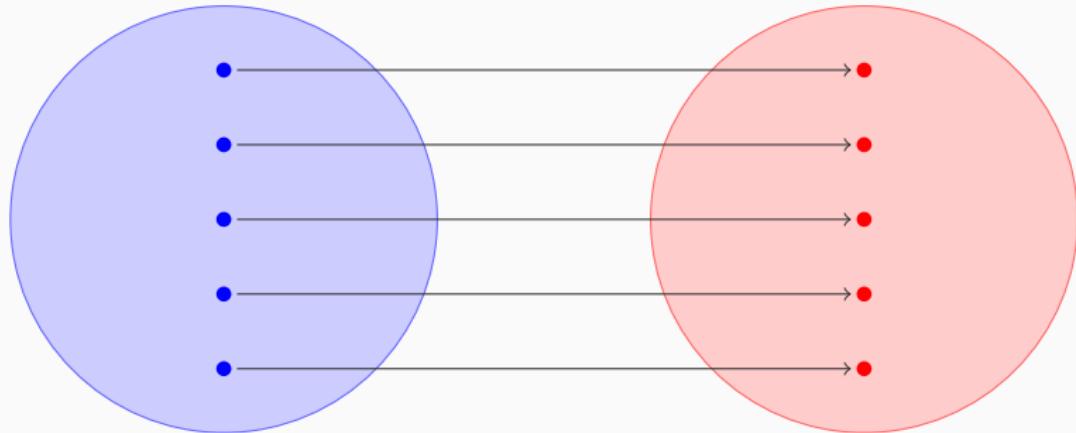
COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$R(x, y) : \Delta \rightarrow \Delta$$



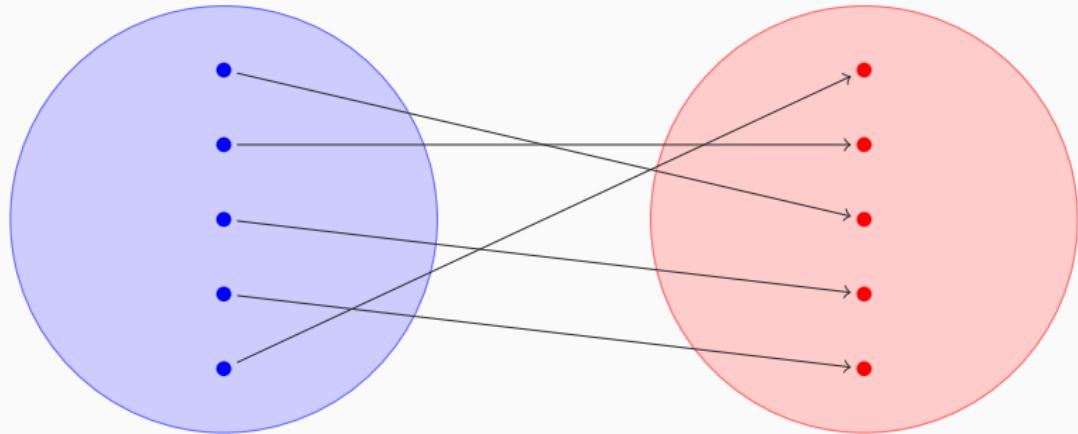
COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$\forall x \exists^{=1} y. R(x, y)$$

 Δ Δ

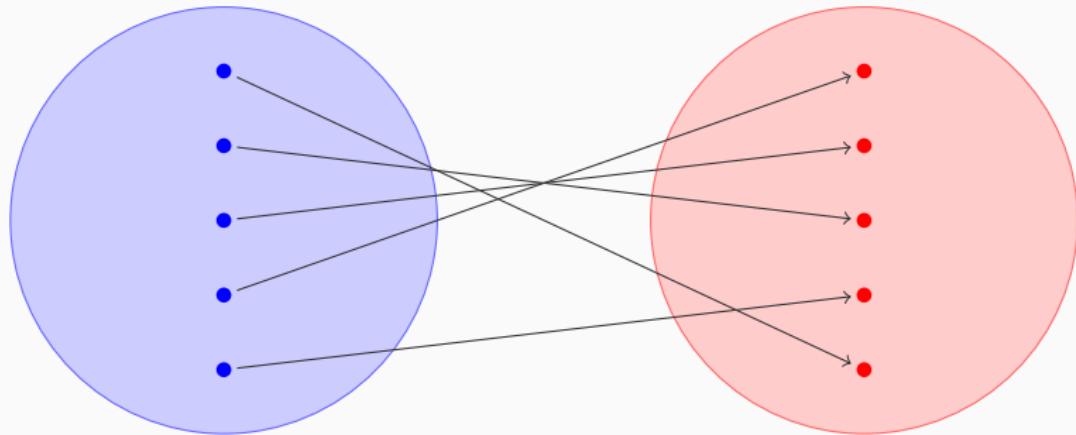
COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$\forall x \exists^{=1} y. R(x, y)$$

 Δ Δ

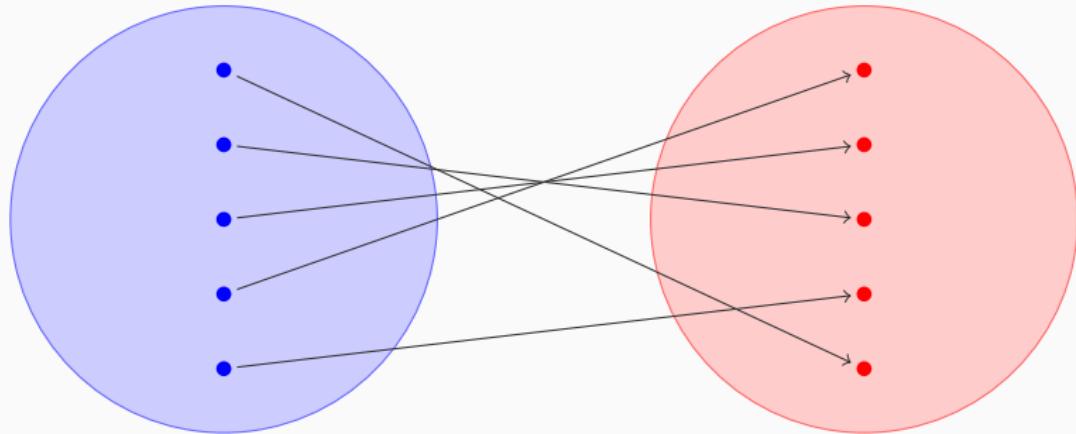
COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$\forall x \exists^{=1} y. R(x, y)$$



COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

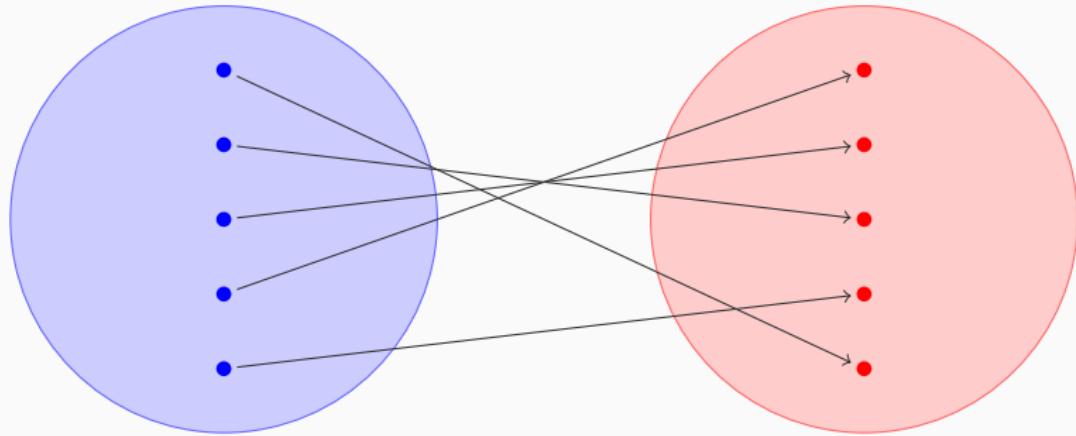
$$\forall x \exists^{=1} y. R(x, y)$$



What is the cardinality of R in all functions ??

COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$\forall x \exists^{=1} y. R(x, y)$$

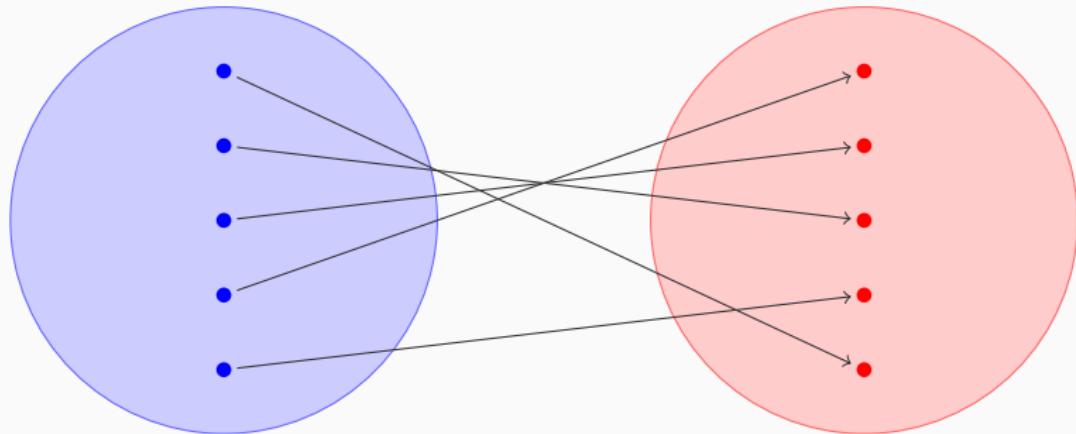


What is the cardinality of R in all functions ??

$$|\Delta|$$

COUNTING QUANTIFIERS: EXPRESSING FUNCTIONALITY

$$\forall x \exists^{=1} y. R(x, y)$$



What is the cardinality of R in all functions ??

$$|\Delta|$$

$$\forall x. \exists^{=1} y. R(x, y) \equiv \forall x. \exists y. R(x, y) \wedge (|R| = |\Delta|)$$

COUNTING QUANTIFIERS⁹

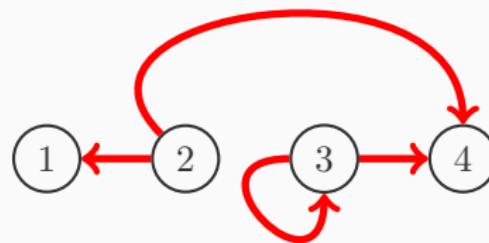
Let Φ be the following C² formula, then it can be written as:

$$\Phi_0 \wedge \bigwedge_{k=1}^q \forall x.(A_k(x) \leftrightarrow \exists^{=m_k} y.R_k(x, y))$$

where Φ_0 is a pure universal formula in FO².

⁹Kuzelka. JAIR 2021, Malhotra and Serafini. AAAI 2022 for this formulation

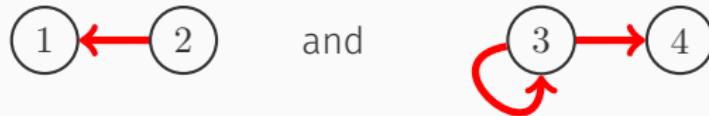
WFOMC BEYOND FIRST-ORDER LOGIC¹⁰



¹⁰Malhotra, Bizzaro and Serafini. 2024 (Under Review at Artificial Intelligence Journal)

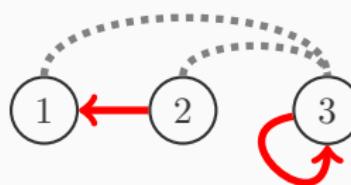
PROJECTIONS ON SUB-DOMAIN

Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as



IN HOW MANY WAYS CAN I MERGE THEM?

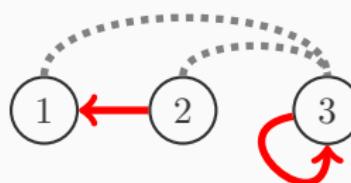
Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as



What is the number of ways to merge these partial interpretations?

IN HOW MANY WAYS CAN I MERGE THEM?

Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as

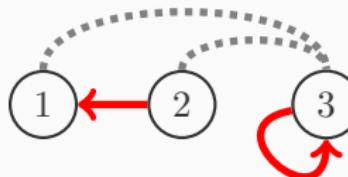


What is the number of ways to merge these partial interpretations?

$$2^4$$

IN HOW MANY WAYS CAN I MERGE UNDER CONSTRAINTS?

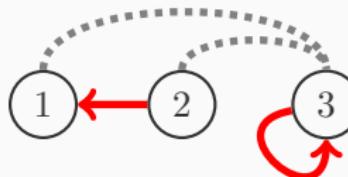
Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as



What is the number of ways to merge these partial interpretations?
Such that there are only directed edges from $[2]$ to $[\bar{2}]$?

IN HOW MANY WAYS CAN I MERGE UNDER CONSTRAINTS?

Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as

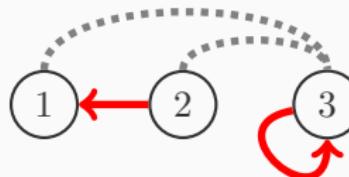


What is the number of ways to merge these partial interpretations?
Such that there are only directed edges from $[2]$ to $[\bar{2}]$?

$$2^2$$

IN HOW MANY WAYS CAN I MERGE UNDER CONSTRAINTS?

Then, $\omega' = \omega \downarrow [2]$ and $\omega'' = \omega \downarrow [\bar{2}]$ are given respectively as



What is the number of ways to merge these partial interpretations?
Such that there are only directed edges from $[2]$ to $[\bar{2}]$?

$$2^2$$

This can be captured by restricting the allowed 2-types between Δ and Δ'

COUNTING BY SPLITTING

How can you compute WFOMC of all the models on $[m]$ and $[\bar{m}]$, such that:

$$C1: \omega \downarrow [m] \models axiom'$$

$$C2: \omega \downarrow [\bar{m}] \models axiom''$$

$$C3: \omega \models \forall x \in [m] \ \forall y \in [\bar{m}]. \Theta(x, y)$$

$$C4: \omega \models \forall xy. \Phi(x, y)$$

COUNTING BY SPLITTING

How can you compute WFOMC of all the models on $[m]$ and $[\bar{m}]$, such that:

$$C1: \omega \downarrow [m] \models axiom'$$

$$C2: \omega \downarrow [\bar{m}] \models axiom''$$

$$C3: \omega \models \forall x \in [m] \ \forall y \in [\bar{m}]. \Theta(x, y)$$

$$C4: \omega \models \forall xy. \Phi(x, y)$$

- $\text{WFOMC}(\Psi', \Psi'') \prod_{i,j \in [u]} r_{ij}^{k'_i k''_j}$, where r_{ij} captures the constraints across them.

-

$$\text{WFOMC}(\Psi_{[m]}, \mathbf{k}) = \sum_{\substack{k'+k''=k \\ |k'|=m}} \text{WFOMC}(\Psi', \mathbf{k}') \text{WFOMC}(\Psi'', \mathbf{k}'') \prod_{i,j \in [u]} r_{ij}^{k'_i k''_j}$$

WHAT CAN I DO WITH THIS RESULT?

	<i>axiom'</i>	<i>axiom''</i>	$\Theta(x, y)$
$DAG(R)$	$\forall xy. \neg R(x, y)$	$DAG(R)$	$\neg R(y, x)$
$Connected(R)$	$Connected(R)$	\top	$\neg R(x, y)$
$Forest(R)$	$Tree(R)$	$Forest(R)$	$\neg R(x, y)$

Table 1: A summary of results using counting by splitting

OPEN PROBLEMS

- Is there a some formal language that completely captures tractable WFOMC?

OPEN PROBLEMS

- Is there a some formal language that completely captures tractable WFOMC?
- Is there an FPRAS? (There is no PTAS (Jaeger 2014. TPLP))

OPEN PROBLEMS

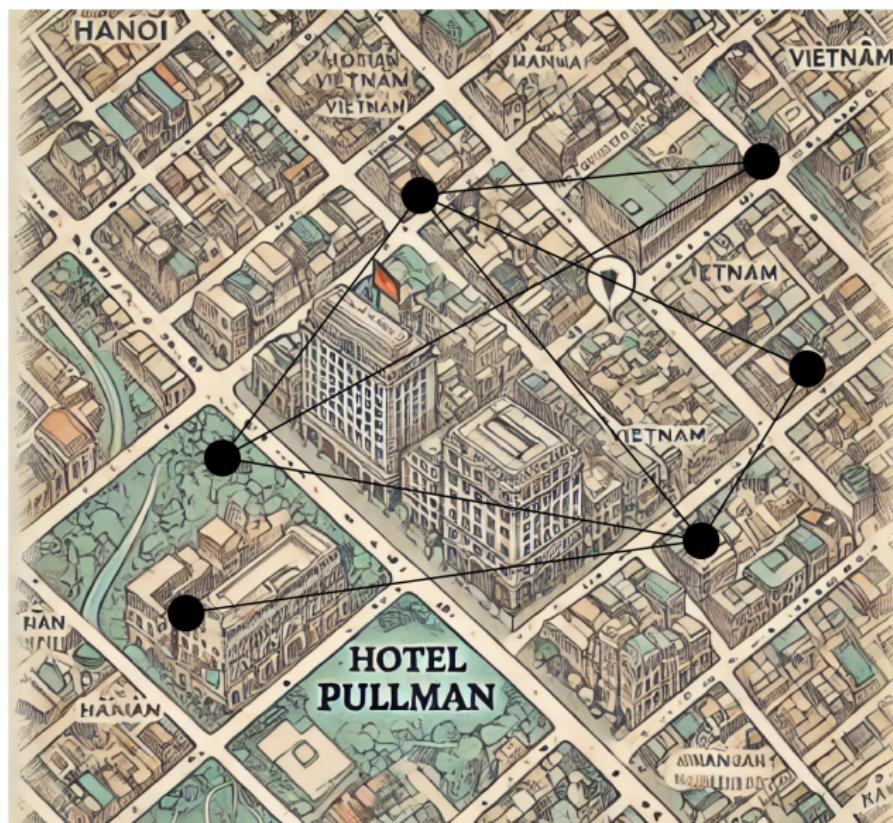
- Is there a some formal language that completely captures tractable WFOMC?
- Is there an FPRAS? (There is no PTAS (Jaeger 2014. TPLP))
- Evidence. Something is known about FPT (Broeck and Darwiche 2013. NeurIPS). But not alot.

OPEN PROBLEMS

- Is there a some formal language that completely captures tractable WFOMC?
- Is there an FPRAS? (There is no PTAS (Jaeger 2014. TPLP))
- Evidence. Something is known about FPT (Broeck and Darwiche 2013. NeurIPS). But not alot.
- Sampling – Amazing recent developments (Wang et al. LICS 2023)

Consistency

A COVID OUTBREAK IN PULLMAN HOTEL!



BACK TO MY GRIM EXAMPLE

A contact-tracing model:

$$w_1 : \text{Covid}(x)$$

$$w_2 : \text{Covid}(x) \wedge \text{Contact}(x, y) \rightarrow \text{Covid}(y)$$

$$\hat{\mathbf{a}} = \operatorname{argmax}_{\mathbf{a}} P_{\Phi}^{(m)}(\omega) \quad (4)$$

A COVID OUTBREAK IN PULLMAN HOTEL!



WHAT DO WE OBSERVE, AND WHAT DO WE LEARN?

How does one usually do parameter estimation?

$$\hat{\mathbf{a}} = \operatorname{argmax}_{\mathbf{a}} P_{\Phi}^{(m)}(\omega) \quad (5)$$

WHAT DO WE OBSERVE, AND WHAT DO WE LEARN?

How does one usually do parameter estimation?

$$\hat{\mathbf{a}} = \operatorname{argmax}_{\mathbf{a}} P_{\Phi}^{(m)}(\omega) \quad (5)$$

What do we actually want to do?

$$\hat{\mathbf{a}} = \operatorname{argmax}_{\mathbf{a}} P_{\Phi}^{(n+m)} \downarrow [n](\omega)$$

where,

$$P^{(n)} \downarrow [m](\omega') = \sum_{\omega \in \Omega^{(n)} : \omega \downarrow [m] = \omega'} P^{(n)}(\omega)$$

CONSISTENCY OF LEARNING AND INFERENCE

Consistency of Inference [Projectivity]¹¹: A model learnt over a domain of size n , basically says nothing quantitative about a domain of size $n + 1$

$$P_{\theta}^{(m)}(q) \neq P_{\theta}^{(n)} \downarrow [m](q)$$

¹¹Shalizi and Rinaldo. 2013. Jaeger and Shulte. StarAI Workshop 2018. Mittal, Bhardwaj, Gogate, Singla. AISTATS 2019.

CONSISTENCY OF LEARNING AND INFERENCE

Consistency of Inference [Projectivity]¹¹: A model learnt over a domain of size n , basically says nothing quantitative about a domain of size $n + 1$

$$P_{\theta}^{(m)}(q) \neq P_{\theta}^{(n)} \downarrow [m](q)$$

Consistency of Learning: Batching or stochastic gradient descent cannot work for MLNs ! as they do not admit consistent parameter estimates

$$\operatorname{argmax}_{\theta} E_{\omega'} [\log P_{\theta}^{(m)}(\omega')] \neq \operatorname{argmax}_{\theta} \log P_{\theta}^{(n)}(\omega)$$

¹¹Shalizi and Rinaldo. 2013. Jaeger and Shulte. StarAI Workshop 2018. Mittal, Bhardwaj, Gogate, Singla. AISTATS 2019.

TRYING TO GET NEAR PROJECTIVITY!

Can we improve the learning procedure to get near-projective models¹²

¹²Chen, Weitkamper, Malhotra. ECML 2024.

(For theory of RMP) Kuzelka, Wang, Davis and Schockaert. AAAI 2018.

TRYING TO GET NEAR PROJECTIVITY!

Can we improve the learning procedure to get near-projective models¹²

$$-\log P_{\Phi}^{(n+m)} \downarrow [n](\omega) \leq -\log P_{\Phi}^{(n)}(\omega) + \log \Delta \quad (6)$$

$$KL(P_{\Phi}^{(n+m)} \downarrow [n] || P_{\Phi}^{(n)}) \leq \log \Delta \quad (7)$$

¹²Chen, Weitkamper, Malhotra. ECML 2024.

(For theory of RMP) Kuzelka, Wang, Davis and Schockaert. AAAI 2018.

StarAI



Tractability



Consistency



BUT WHAT ARE PROJECTIVE MODELS?

BUT WHAT ARE PROJECTIVE MODELS?

Erdős–Rényi random graph:

$$\Pr(G(n+m, p) = G) = p^{|E|} (1-p)^{\binom{n+m}{2} - |E|}$$

BUT WHAT ARE PROJECTIVE MODELS?

Erdős–Rényi random graph:

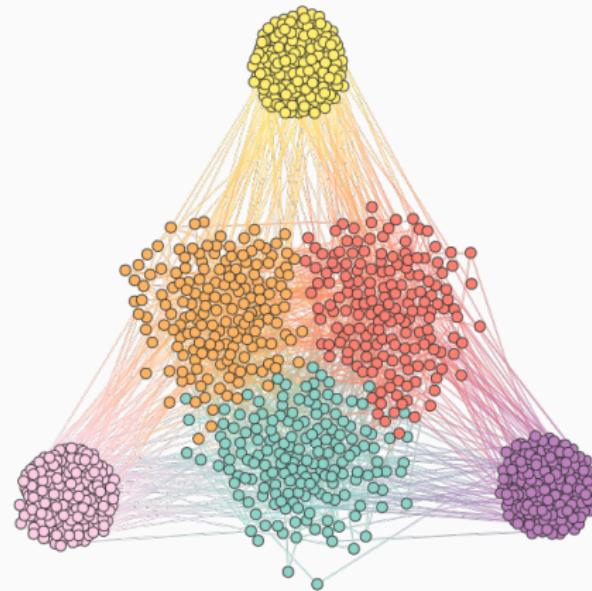
$$\Pr(G(n+m, p) = G) = p^{|E|} (1-p)^{\binom{n+m}{2} - |E|}$$

$$\Pr(G(n+m, p) \downarrow [n] = G') = p^{|E'|} (1-p)^{\binom{n}{2} - |E'|} \quad (8)$$

$$= \Pr(G(n, p) = G'). \quad (9)$$

CAN WE EXPRESS ANY PROJECTIVE MODELS IN MLNs?

"A Markov Logic Network in the Two Variable Fragment is projective if and only if it represents a stochastic block structure¹³"



¹³Malhotra and Serafini. ECML 2022

PROBABILITY PARAMETERS

- p_i is the probability of an arbitrary domain constant realising the i^{th} 1-type

$$p_i := P(i(x))$$

PROBABILITY PARAMETERS

- p_i is the probability of an arbitrary domain constant realising the i^{th} 1-type

$$p_i := P(i(x))$$

- w_{ijl} is the conditional probability of an arbitrary pair of domain constants to realise the l^{th} 2-table, given they realise the i^{th} and the j^{th} 1-type.

$$w_{ijl} = P(l(x, y) | i(x) \wedge j(y))$$

“WHO SHAVES WHO IN KR?”

Assuming **Shaves** to be irreflexive and symmetric, we have the following non-zero $\{p_i, w_{ijl}\}$

$$p_1 = P(\text{KR(a)})$$

$$p_2 = P(\neg \text{KR(a)})$$

$$w_{111} = P(\text{Shaves(a, b)} | \text{KR(a)}, \text{KR(b)})$$

$$w_{112} = P(\neg \text{Shaves(a, b)} | \text{KR(a)}, \text{KR(b)})$$

$$w_{121} = P(\text{Shaves(a, b)} | \text{KR(a)}, \neg \text{KR(b)})$$

$$w_{122} = P(\neg \text{Shaves(a, b)} | \text{KR(a)}, \neg \text{KR(b)})$$

$$w_{221} = P(\text{Shaves(a, b)} | \neg \text{KR(a)}, \neg \text{KR(b)})$$

$$w_{222} = P(\neg \text{Shaves(a, b)} | \neg \text{KR(a)}, \neg \text{KR(b)})$$

THE RELATIONAL BLOCK MODEL

Relational Block Model :

$$P(\mathbf{X} = \mathbf{x}) := \prod_{q=1}^n p_{x_q} = \prod_{i=1}^u p_{x_i}^{k_i}$$

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) &:= \prod_{1 \leq q < r \leq n} w_{x_q x_r} y_{qr} \\ &= \prod_{1 \leq i \leq j \leq u} \prod_{1 \leq l \leq b} (w_{ijl})^{h_l^{ij}} \end{aligned}$$

Any projective MLN in the two variable fragment reduces to an RBM

WHAT IS THE MOST GENERAL CLASS OF PROJECTIVE MODELS ON GRAPHS?

¹⁴László Lovász and Aldous-Hoover-Kallenberg

WHAT IS THE MOST GENERAL CLASS OF PROJECTIVE MODELS ON GRAPHS?

A **graphon**¹⁴ is a symmetric, measurable function

$W : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that:

- $W(x, y) = W(y, x)$ for all $x, y \in [0, 1]$ (symmetry),
- $W(x, y)$ gives the probability of an edge between points x and y .

¹⁴László Lovász and Aldous-Hoover-Kallenberg

WHAT IS THE MOST GENERAL CLASS OF PROJECTIVE MODELS ON GRAPHS?

A **graphon**¹⁴ is a symmetric, measurable function

$W : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that:

- $W(x, y) = W(y, x)$ for all $x, y \in [0, 1]$ (symmetry),
- $W(x, y)$ gives the probability of an edge between points x and y .

Sampling a graph from a graphon:

1. Sample n points u_1, u_2, \dots, u_n uniformly from $[0, 1]$.
2. For each pair (i, j) with $i \neq j$, place an edge with probability $W(u_i, u_j)$.

¹⁴László Lovász and Aldous-Hoover-Kallenberg

StarAI

ooooooooo

Tractability

oooooooooooooooooooooooooooooooooooo

Consistency

oooooooooooooo●ooo

WHAT DO ALL THESE MODELS HAVE IN COMMON?

WHAT DO ALL THESE MODELS HAVE IN COMMON?

- Exchangeable Distributions:

$$P((X_{ij})) = P((X_{\sigma(i)\sigma(j)}))$$

WHAT DO ALL THESE MODELS HAVE IN COMMON?

- Exchangeable Distributions:

$$P((X_{ij})) = P((X_{\sigma(i)\sigma(j)}))$$

- X_{ij} are edges for a ER graph or a graphon
- X_{ij} are 2-types for FO^2
- Can be generalized to exchangeable random arrays like X_{ijl}

ALDOUS-HOOVER-KALLENBERG

Theorem 1 (Aldous-Hoover representation of jointly exchangeable matrices (Aldous, 1981; Hoover, 1979)). A random 2-array $(X_{ij})_{i,j \in \mathbb{N}}$ is jointly exchangeable if there exists a function

$f : [0, 1] \times [0, 1]^2 \times [0, 1] \rightarrow E$ such that

$$(X_{ij}) = (f(U, U_i, U_j, U_{ij})) ,$$

where $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i,j > i \in \mathbb{N}}$ with $U_{ij} = U_{ji}$ are a sequence and matrix, respectively, of i.i.d. Uniform[0, 1] random variables.

OPEN PROBLEMS

- How to represent and learn expressive AHK models?

OPEN PROBLEMS

- How to represent and learn expressive AHK models?
- How to encode expert Knowledge into AHK models?

StarAI

ooooooooo

Tractability

oo

Consistency

oooooooooooooooooooo●

Thank You!