WWU
MÜNSTER

Lifted Inference

Workshop UzL&WWU

Tanya Braun, University of Münster

living.knowledge

Research Group Data Science
Computer Science Department

Agenda

Workshop

- Topic:
  Lifted inference in its broadest sense

- Goal:

  Getting to know each other and our current research

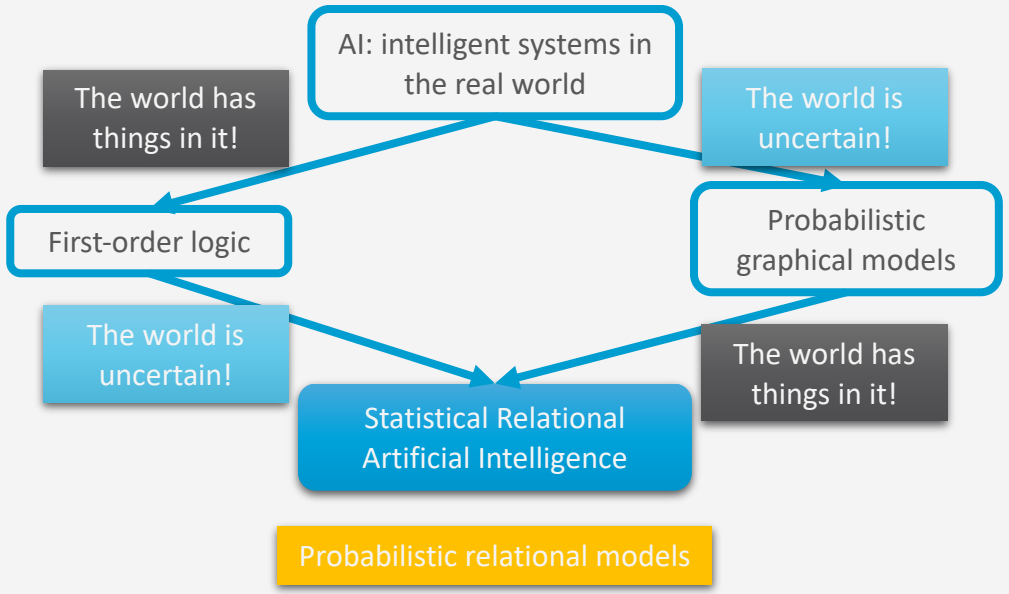
  Exchange of ideas and knowledge

Schedule

- 13.00: begin
  - Talks of ~20mins by participants
    - Proposed order:
      - Tanya [going on]
      - Marcel
      - Malte
      - Sagad
      - Mattis
- 15.00: coffee break and start of discussion
- 17.30: tentative end

Overview

The Power of Indistinguishability



Statistical Relational Artificial Intelligence (StaRAI)

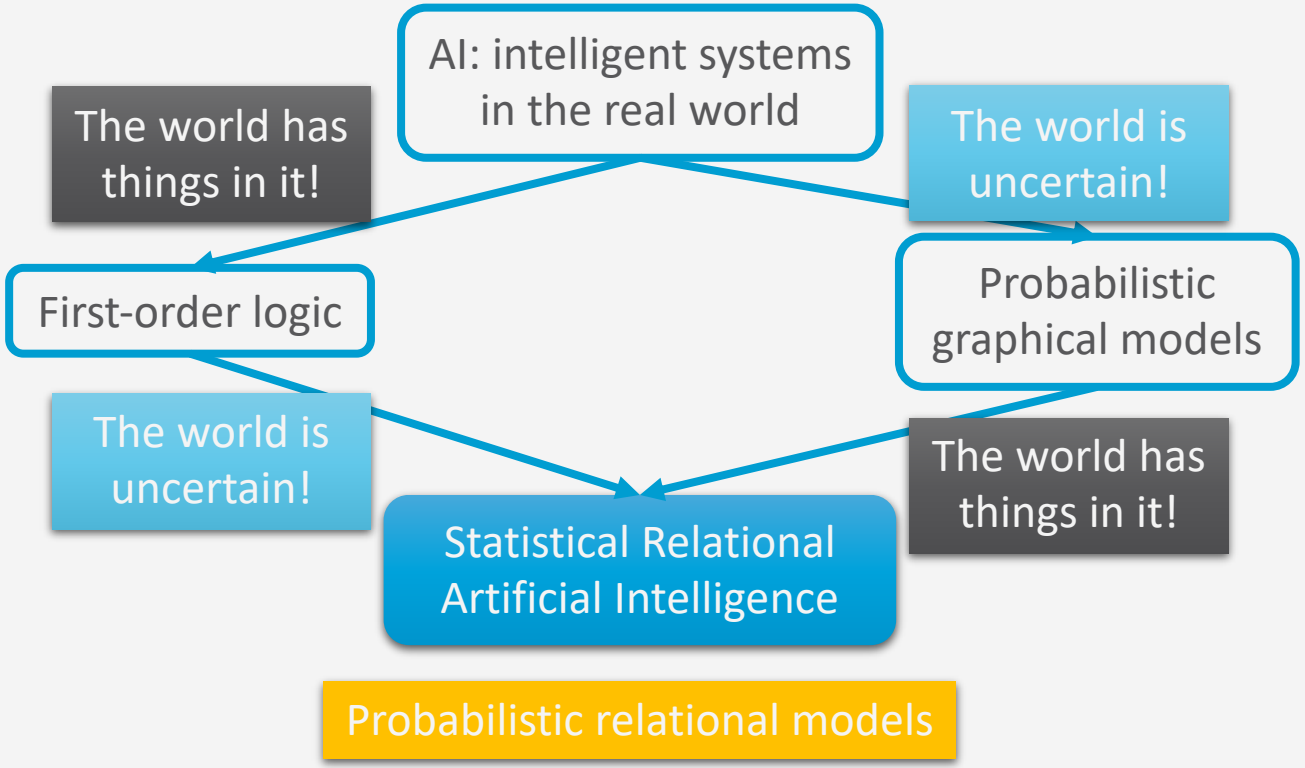
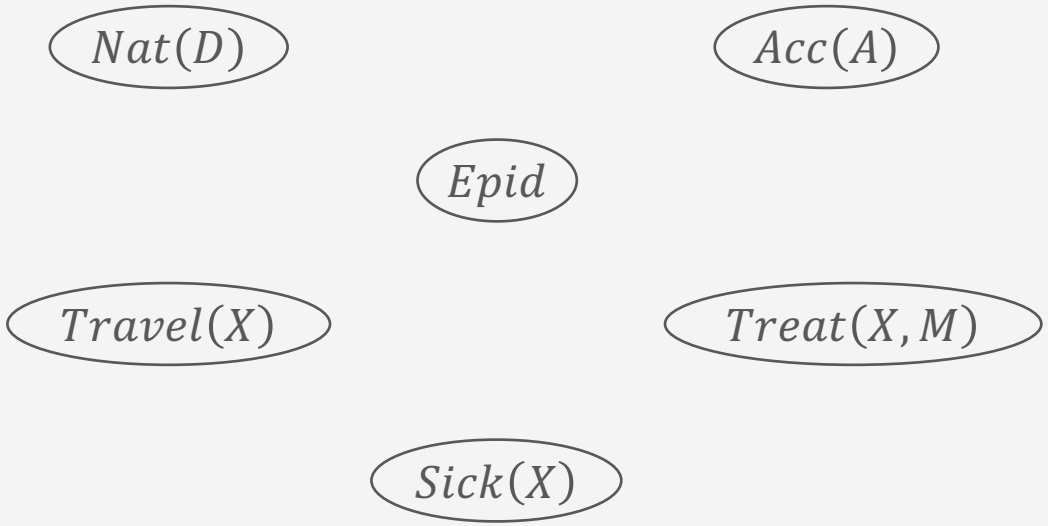


Figure based on Stuart Russell

Application: Epidemics

- 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
    - $ran\big(Travel(X)\big) = \{true, false\}$
  - Represent sets of *indistinguishable* random variables

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

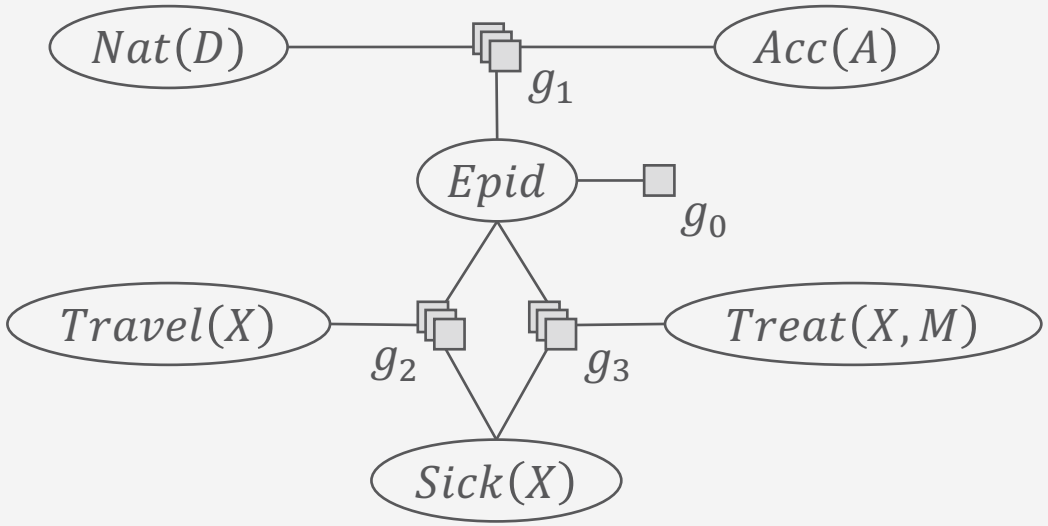


Encoding the Joint Distribution: Factorisation

- 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
$true$	$false$	$true$	6
$true$	$true$	$false$	2
$true$	$true$	$true$	9

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

$reat(X, M)$

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]

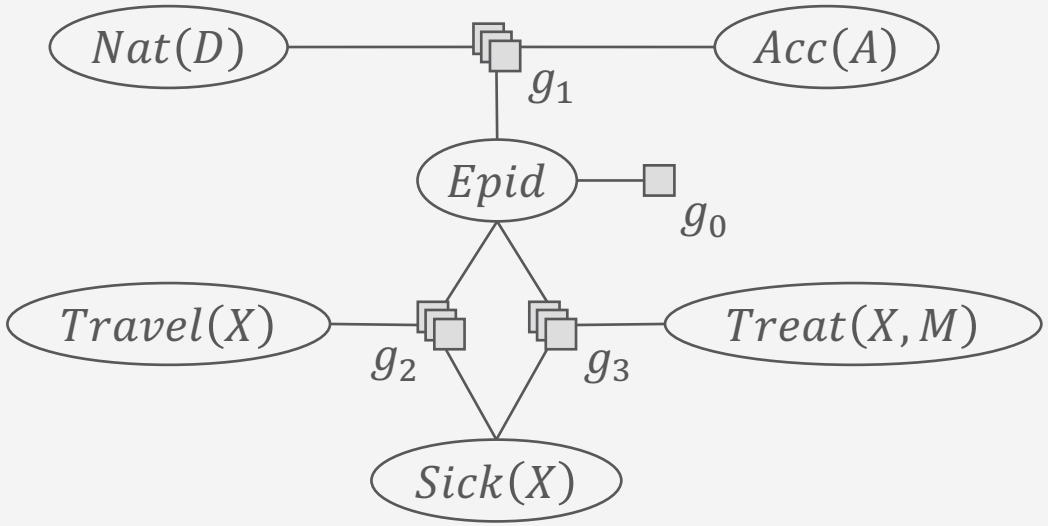
$$P_G = \frac{1}{Z} \prod_{f \in gr(G)} f$$

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

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

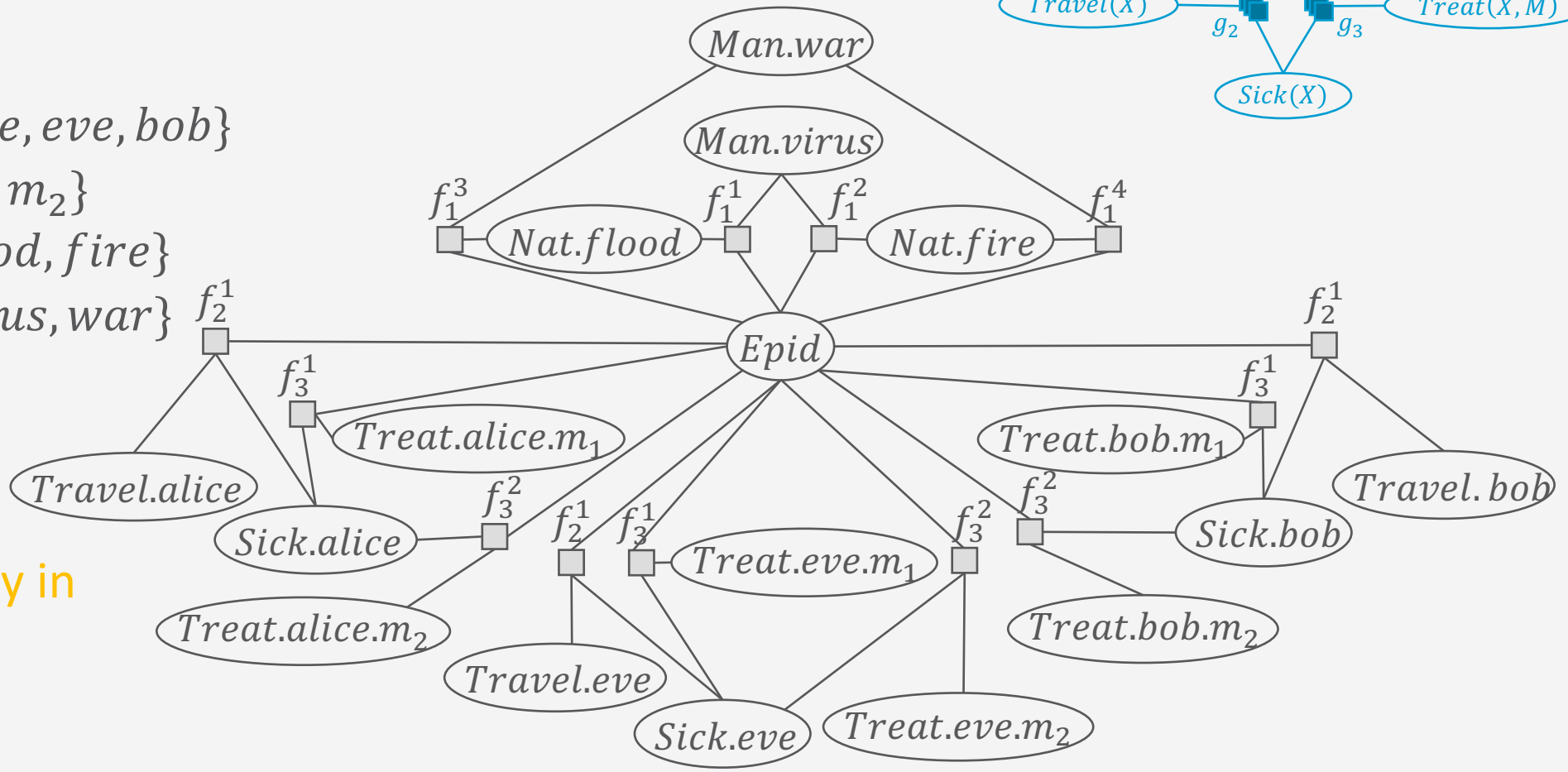
Sparse encoding of joint distribution

$3 \cdot 2^3 = 24$ entries in 3 parfactors, 6 PRVs



Grounded Model

- Given domains
  - $dom(X) = \{alice, eve, bob\}$
  - $dom(M) = \{m_1, m_2\}$
  - $dom(D) = \{flood, fire\}$
  - $dom(W) = \{virus, war\}$



- Indistinguishability in
  - Graph structure
  - Factors

Probabilistic Relational Models and Variants

- Parfactors Models
  [Poole 03, Taghipour et al. 13, B & Möller 16-19, Gehrke, B & Möller 18-19]

- Markov Logic Networks (MLNs) [Richardson & Domingos 06]
  - Use logical formulas to specify potential functions

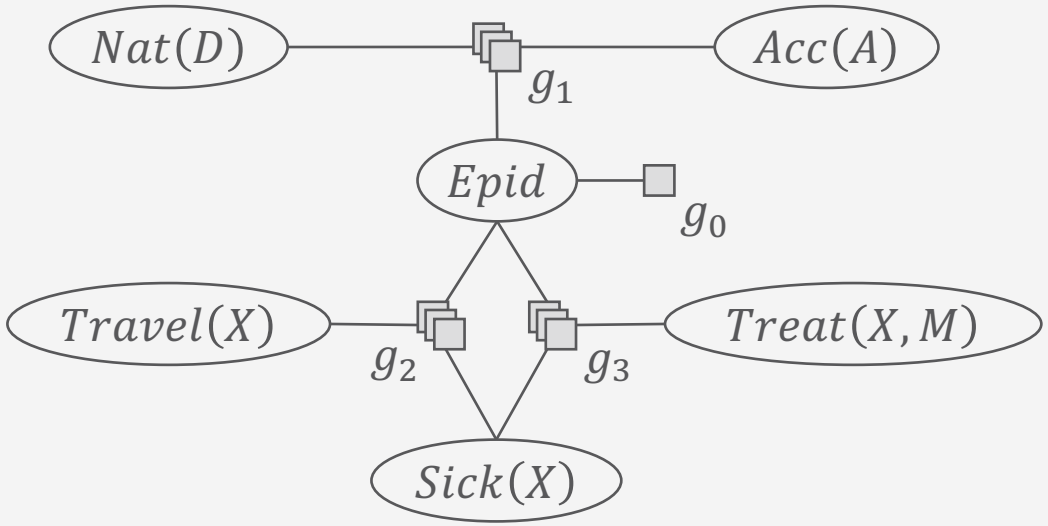
- Probabilistic Soft Logic (PSL) [Bach et al. 17]
  - Use density functions to specify potential functions

- Based on grounding semantics [Sato 95, Fuhr 95]

Reasoning on Probabilistic Relational Models

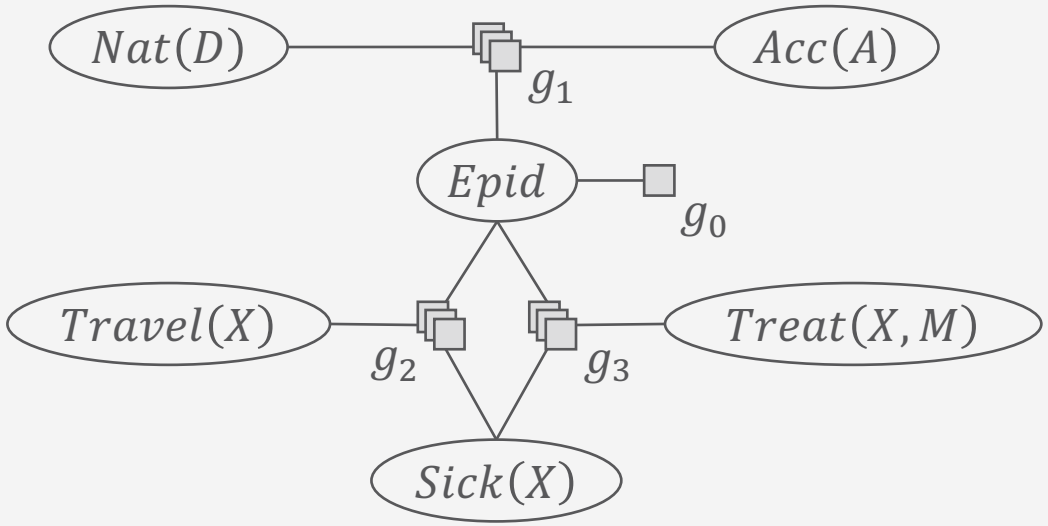
- Inference task: query answering (QA)
- Queries:
  - Marginal distribution
    - $P(Sick(eve))$
    - $P(Travel(eve,)\ Treat(eve, m_1))$
  - Conditional distribution
    - $P(Sick(eve)|Epid)$
    - $P(Epid|Sick(eve) = true)$
  - Assignment queries: $\arg\max_{\boldsymbol{a} \in ran(\boldsymbol{A})} P(\boldsymbol{a}|\boldsymbol{e})$
    - MPE: $\boldsymbol{A} = rv(\boldsymbol{G}) \setminus rv(\boldsymbol{e})$
    - MAP: $\boldsymbol{A} \subseteq rv(\boldsymbol{G}) \setminus rv(\boldsymbol{e})$
      - What is not in $\boldsymbol{A}$ needs to be summed out

Goal: Avoid groundings!
→ *lifted* inference



QA: Lifted Variable Elimination (LVE)

- Eliminate all variables not appearing in query
- Lifted summing out
  - Sum out *representative* instance as in propositional variable elimination
  - Exponentiate result for indistinguishable 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



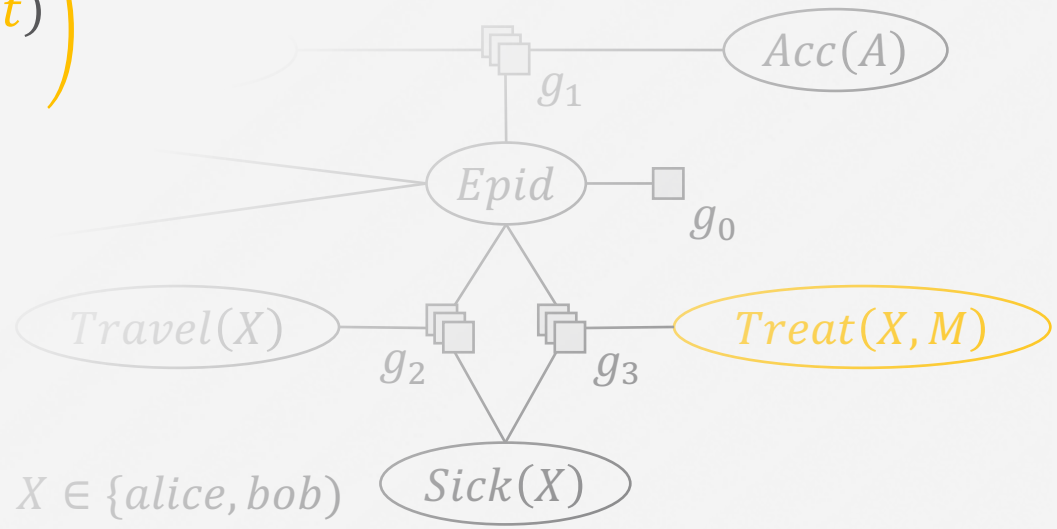
[Poole 03, de Salvo Braz et al. 05, 06, Milch et al. 08, Taghipour et al. 13, 13a]

LVE in Detail: Lifted Summing Out

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

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

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



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
- An 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)

Indistinguishable Evidence and Query Terms

Evidence

- Observations for instances of a PRV
  - One of the range values
  - Not available
- Treat as groups per observation
  - Shatter model on the groups
- Example: 10 instances observed true

$Sick(X^T)$	g_e^T
$false$	0
$true$	1

$$dom(X^T) = \{x_1, \dots, x_{10}\}$$
$$dom(X) = \{x_{11}, \dots, x_n\}$$

Query Terms

- Indistinguishable instances in query:
  - $P\big(Sick(alice), Sick(eve), Sick(bob)\big)$
  - Result will have local symmetries, e.g., 2 false and 1 true maps to potential of 2
- Parameterised query: $P\big(Sick(X)\big)$
- Use standard LVE
  - Count conversion yields wanted result

$\#_X[Sick(X)]$	g
[0,3]	1
[1,2]	2
[2,1]	3
[3,0]	4

Indistinguishability in Other Forms

- Sequential / temporal formalisms
  [Marcel's diss]
  - Sequential parfactor graphs
  - Sequential inference over interfaces: filtering, prediction, hindsight
- Decision-theoretic models
  [Marcel's diss; ongoing research Marcel & me]
  - Parameterised decisions and utilities
  - Explanation & exploration
  - Lifted multi-agent decision making
    - Partitioned agent set in DecPOMDPs
- Continuous formalisms
  [Mattis' diss]
  - Parameterised Gaussian Bayesian networks
  - Inference over lifted full joint over means and covariance matrix
- Hybrid formalisms
  [high interest by Mattis & me to combine our research]
  - Discrete and continuous variables
  - Not a lot exists

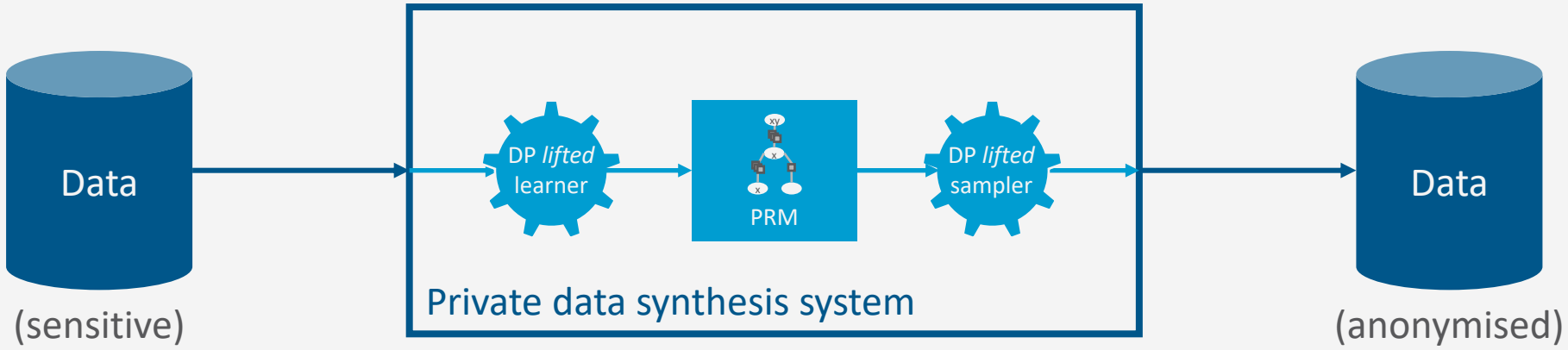
Indistinguishability in Other Forms

- Causal formalisms
  [part of Malte's focus in his diss]
  - Relational causal models
  - Lifted do-calculus

- Finding and encoding indistinguishability
  - Encoding a propositional graph as a parfactor graphs
    [part of Malte's focus]
  - Finding or constructing indistinguishability in not symmetric models
    [part of Sagad's focus]

- Lifting for privacy
  - Temporal inference and online learning
    [part of Marcel's focus]
  - Learning a relational model for privacy-preserving data synthesis
    [upcoming project proposal by Esfandiar (IT Security, UzL) & me]

What else is there to do? – Oh, so much…
- Generalising lifting operators
- More robust learning algorithms
- Ethical behaviour
- Explainability
- …

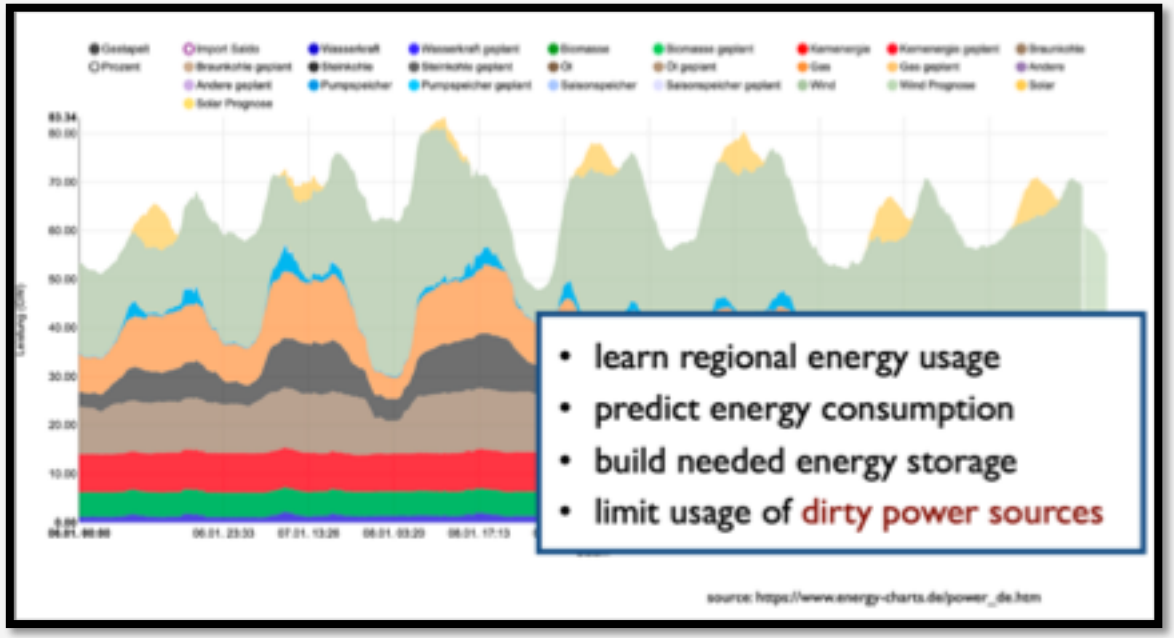


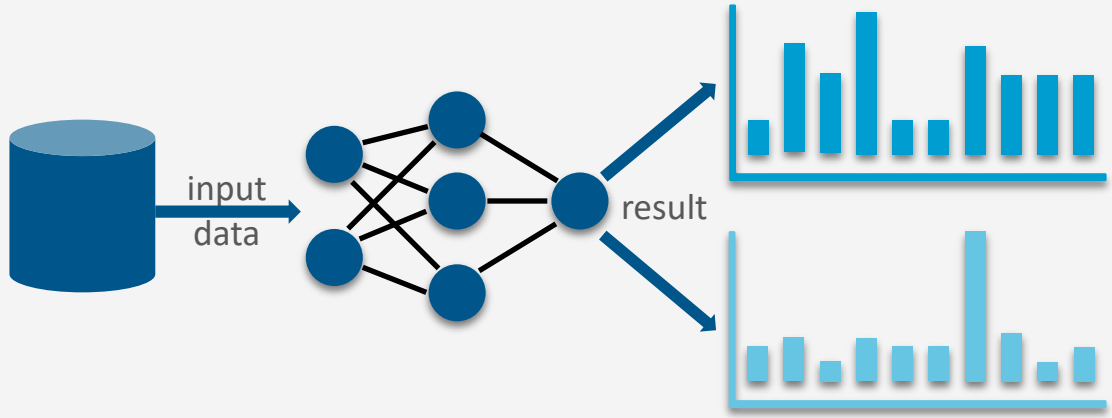
PRIVLIFT

Private Data Synthesis Via Lifting

Setting

- Ongoing collection of huge amounts of data
- Huge potential for research and development
  - Requires data to be published
- Fact: Data often includes sensitive data, e.g.,
  - Medical data
  - Energy-consumption data
- Problem: Training models on this data may potentially reveal sensitive information
  - Especially if the model provides further information such as a confidence score or a distribution over possible classifications
- Solution?





Solution Approaches

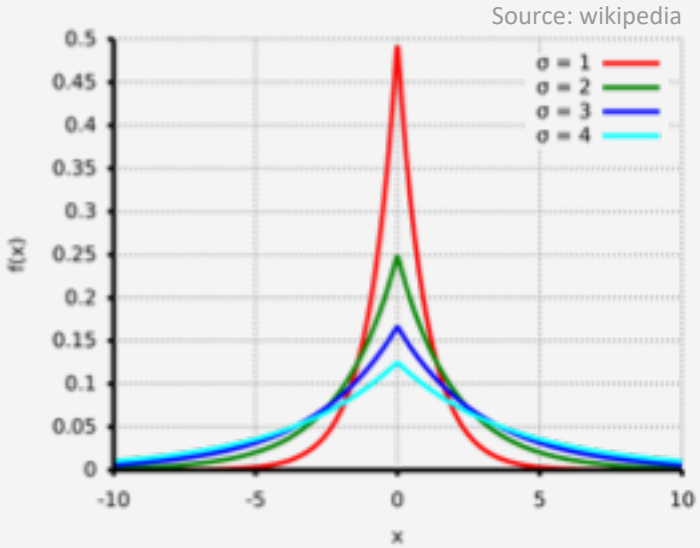
- Simple anonymisation techniques
  - No direct identifiers
  - Ranges of values
- Make properties hold
  - E.g., k-anonymity
    - Each combination of attribute values that occurs occurs at least k times
    - Problem: attacker might have or learn of new information that then allows to identify a person (→ Netflix & IMDB case)
  - E.g., differential privacy + PGMs
    - Difference in one entry between two data sets does not have an effect on the output

Name	Age	Gender	Semester	Grade	Minor
*	17-20	*	1	1.3	Math
*	17-20	*	1	2.0	Literature
*	17-20	*	1	1.7	Philosophy
*	17-20	*	1	3.7	CS
*	17-20	*	1	1.0	CS
*	17-20	*	3	1.3	History
*	21-25	*	3	2.3	Math
*	21-25	*	3	3.0	CS
*	17-20	*	3	failed	CS
*	17-20	*	3	1.7	Literature
*	21-25	*	3	1.0	Physics
*	21-25	*	5	3.3	Math
*	21-25	*	5	1.7	CS
*	21-25	*	5	failed	History
*	17-20	*	5	2.7	Literature
*	21-25	*	5	3.0	Math
*	17-20	*	5	1.0	Physics

Differential Privacy

- Idea: Restrict influence of single data points
  - Drawing conclusions about specific data points limited
- Formal idea: Attacker cannot see any difference between two data sets D, D', which only differ in one data point
  - Data point replaced or added / removed
- One approach for counting queries
  1. Compute query result $q(D)$
  2. Add Laplace noise $Lap(0, \sigma)$ to $q(D)$
     - Sample value v from $Lap(0, \sigma)$ distribution and calculate $q(D) + v$
       - I.e., $Lap(q(D), \sigma)$
     - Mean 0
     - Scaling σ

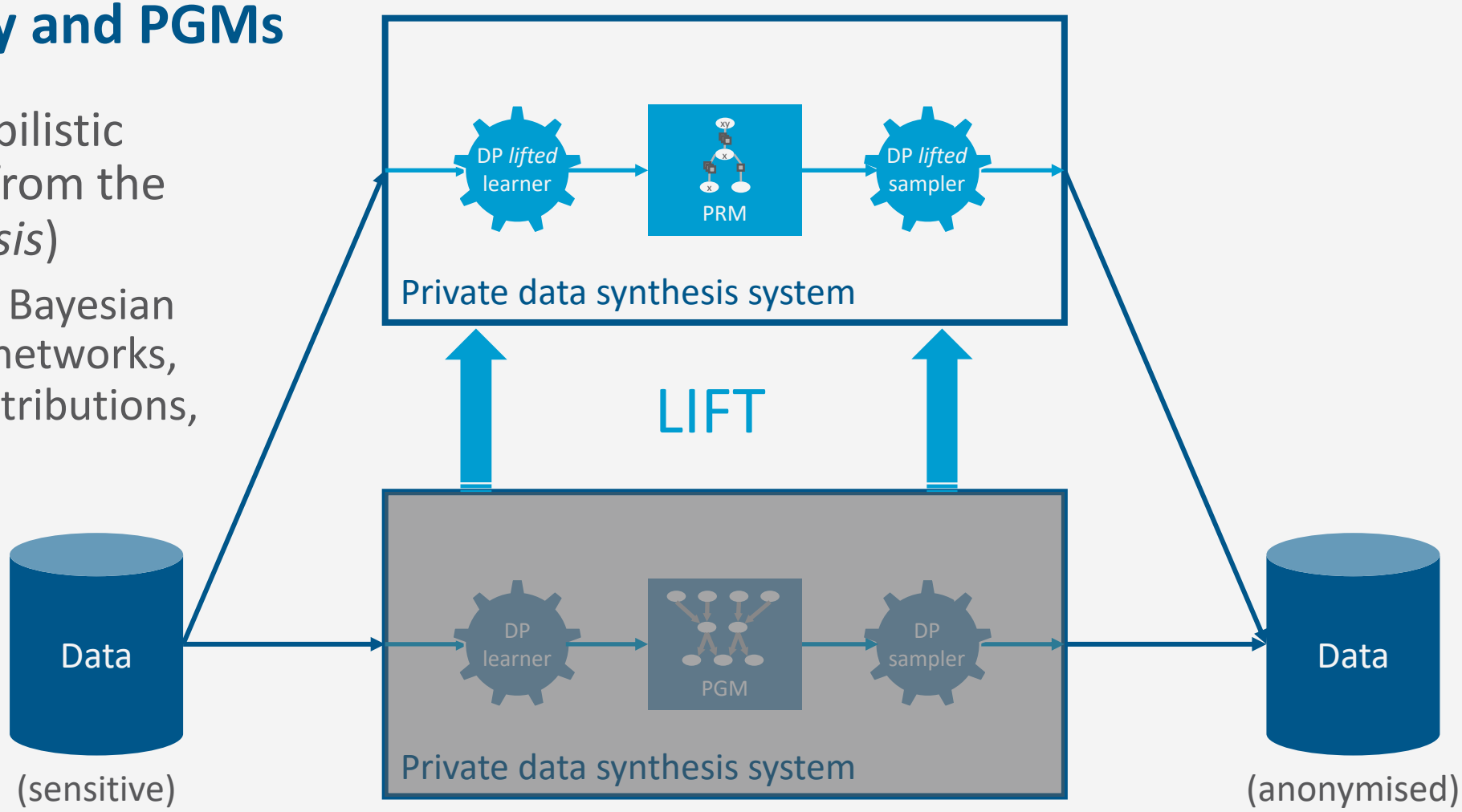
Source: wikipedia



Laplace distribution for values in $[-10,10]$

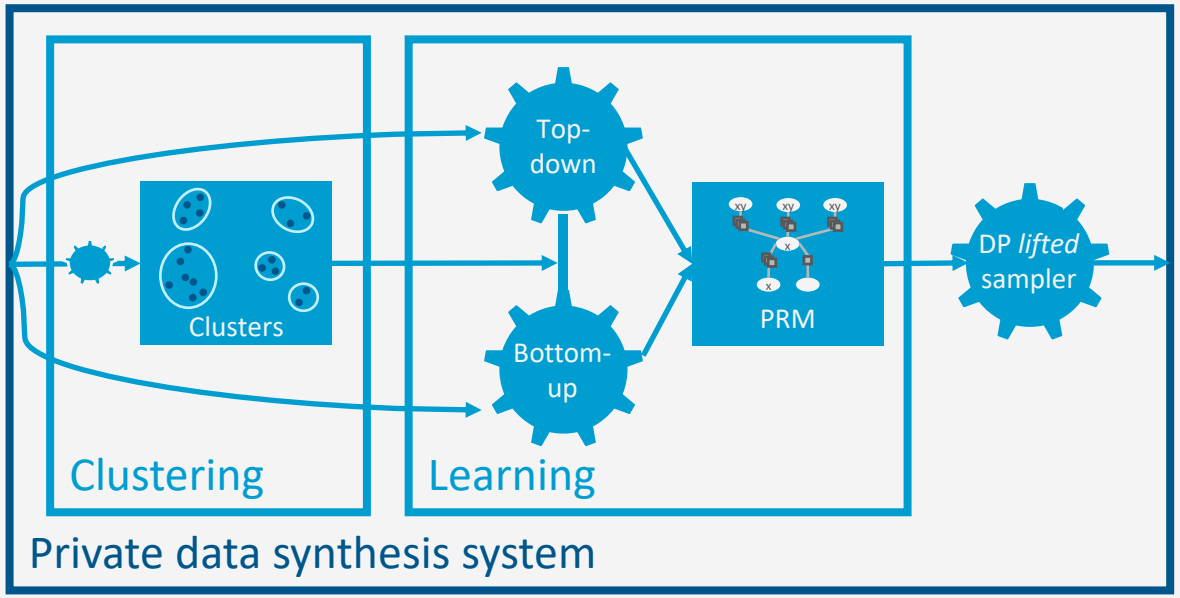
Differential Privacy and PGMs

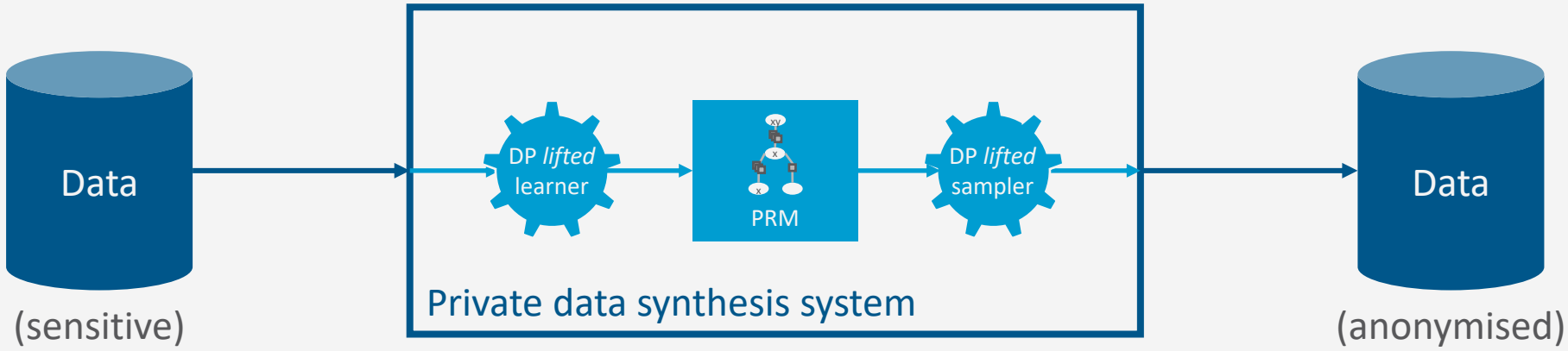
- Idea: Learn a probabilistic model and sample from the model (*data synthesis*)
  - Existing work using Bayesian networks, Markov networks, sets of marginal distributions, …
- Cannot handle relational data very well
  - Has to learn special cases (against DP)



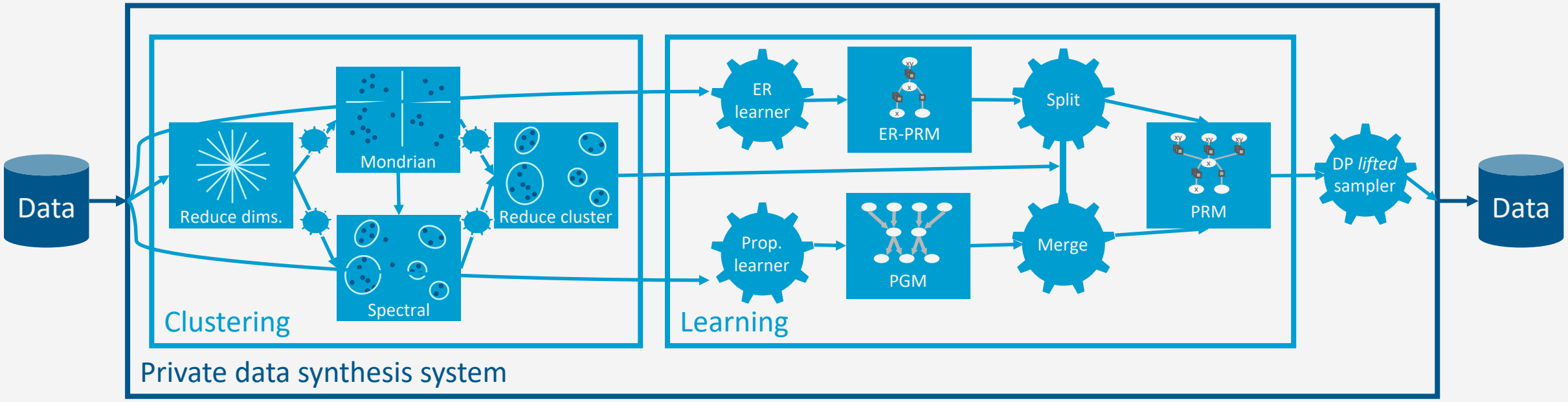
PrivLift

- Idea: Learn a *PRM* and sample from the model in a lifted way
  - Learning how?
    - Top-down
    - Bottom-up
  - Requires clustering
    - DP!





PrivLift



- Clustering: top-down / bottom-up; reduce dimensions for privacy, clusters for efficiency
- Top-down: ER diagram to PRM with top domains → split (Mondrian; uncertain evidence)
- Bottom-up: propositional PGM → merge (spectral; colour passing)

Summary

- Goal: publish sensitive data in an anonymised fashion
- Private data synthesis via lifting
  - Learn a PRM in a private way
    - Requires clustering to split or merge
    - Top-down or bottom-up
      - Learn a coarse model and split
      - Learn a detailed model and merge
  - Sample from the PRM in a private way
    - Sampling not expensive in terms of privacy
    - Lifted sampling → tractability under certain conditions