Dynamic StarAI

Answering Continuous Queries

Tutorial at KI 2019

Tanya Braun, Marcel Gehrke, Ralf Möller
Universität zu Lübeck

UNIVERSITÄT ZU LÜBECK

Agenda: Dynamic Models and Statistical Relational AI

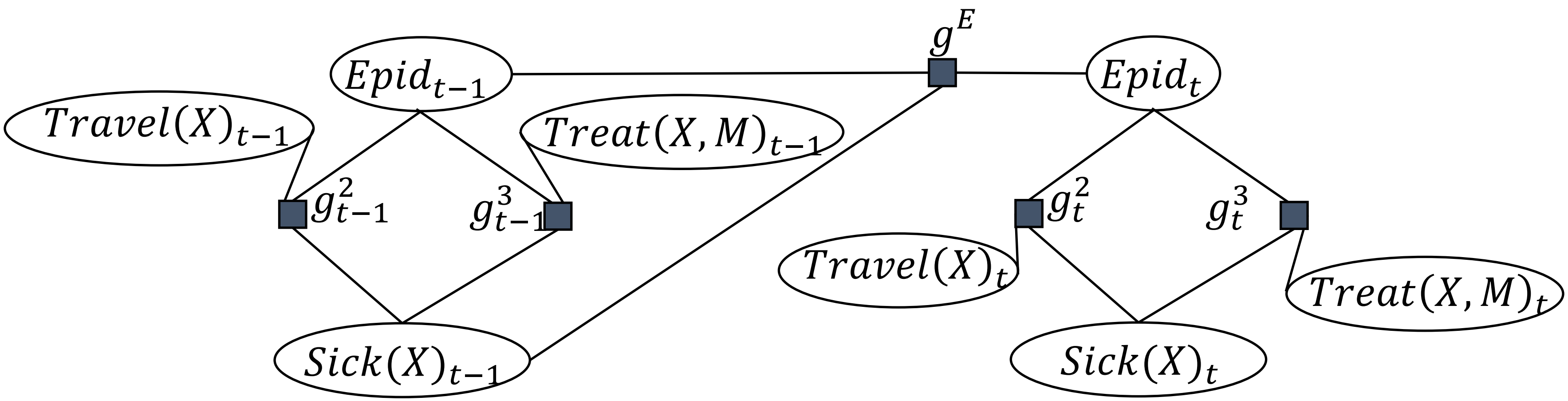
- Probabilistic relational models (PRMs) (Ralf)
- Answering static queries (Tanya)
- **Answering continuous queries (Marcel)**
  - **Lifted Dynamic Junction Tree Algorithm (LDJT)**
  - **Relational interfaces**
  - **Taming reasoning w.r.t. lots of evidence over time**
- Take home messages (Ralf)
  - LJT and LDJT research
    relevant for all variants of PRMs

Goal:
Overview of
central ideas

Lifted: Dynamic Model

Gehrke et al. (2018)

- Marginal distribution query: $P\left(A_{\pi}^{i} \,\middle|\, E_{0:t}\right)$ w.r.t. the model:
  - Hindsight: $\pi < t$ (was there an epidemic t $- \pi$ days ago?)
  - Filtering: $\pi = t$ (is there an currently an epidemic?)
  - Prediction: $\pi > t$ (is there an epidemic in $\pi - t$ days?),



UNIVERSITÄT ZU LÜBECK

Lifted Dynamic Junction Tree Algorithm: LDJT

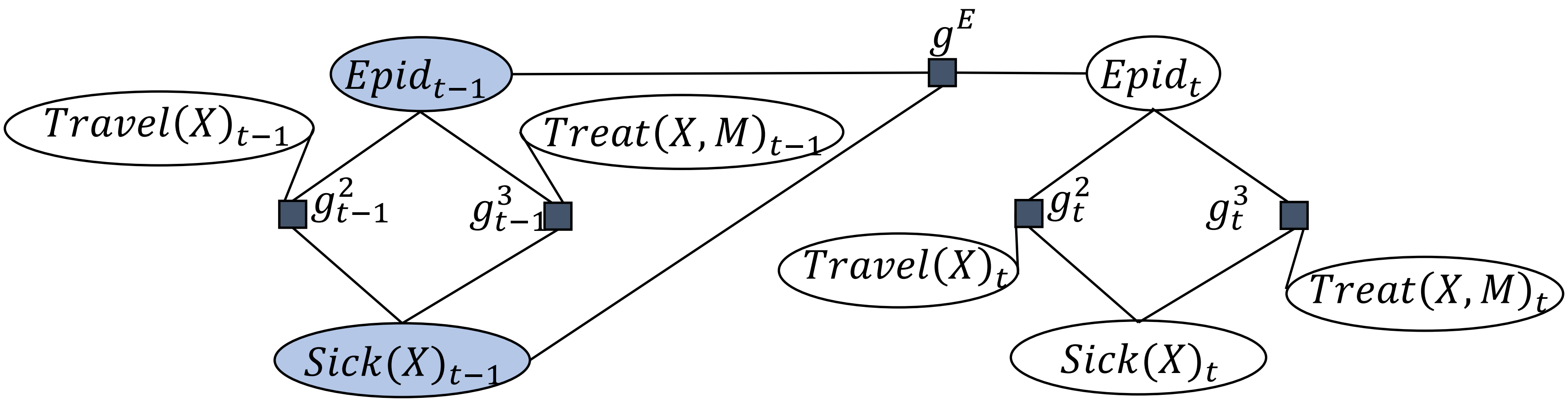
Gehrke et al. (2018)

- Input
  - Temporal model G
  - Evidence $\boldsymbol{E}$
  - Queries $\boldsymbol{Q}$
- Algorithm
  1. Identify interface variables
  2. Build FO jtree structures J for G
  3. Instantiate J_t
  4. Restore state description of interface variables from m_{t-1}
  5. Enter evidence $\boldsymbol{E}_t$ into J_t
  6. Pass messages in J_t
  7. Answer queries $\boldsymbol{Q_t}$
  8. Store state description of interface variables in m_t
  9. Proceed to next time step (step 3)

LDJT: Identify Interface Variables

Gehrke et al. (2018)

- Use temporal conditional independences to perform inference on smaller model (Murphy (2002))
- $\boldsymbol{I}_{t-1} = \{A^i_{t-1} \mid \exists\, \phi(\mathcal{A})_{|C} \in G : A^i_{t-1} \in \mathcal{A} \;\; \wedge A^j_{t-1} \in \mathcal{A}\}$
- Set of interface variable $\boldsymbol{I}_{t-1}$ consists of all PRVs from time slice $t-1$ that occur in a parfactor with PRVs from time slice t

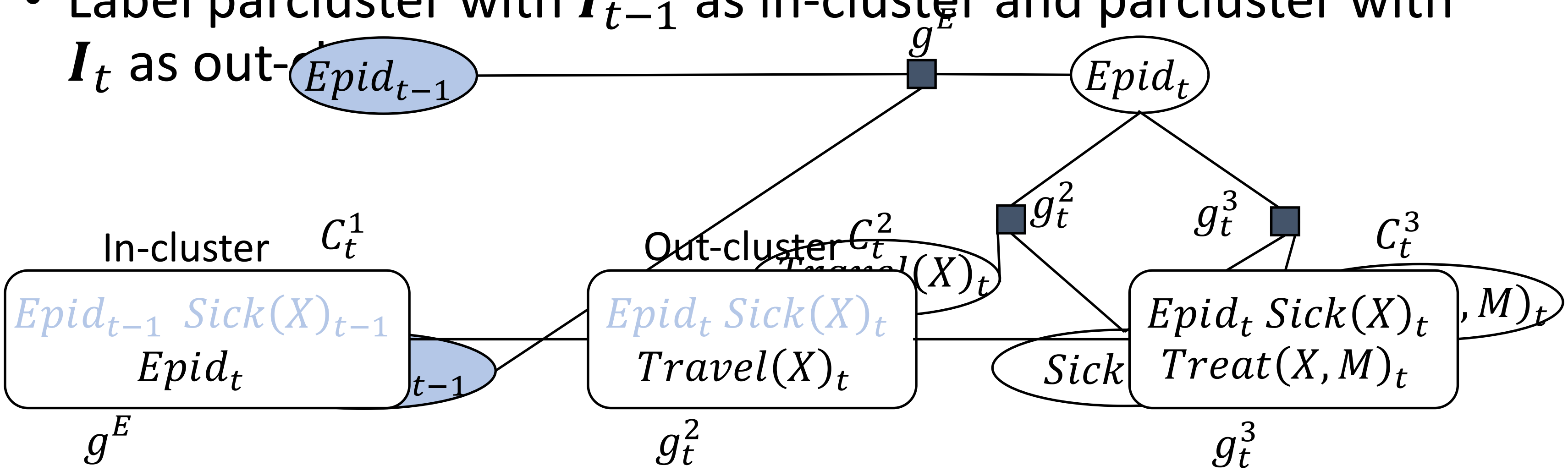


UNIVERSITÄT ZU LÜBECK

LDJT: Construct FO jtree Structure

Gehrke et al. (2018)

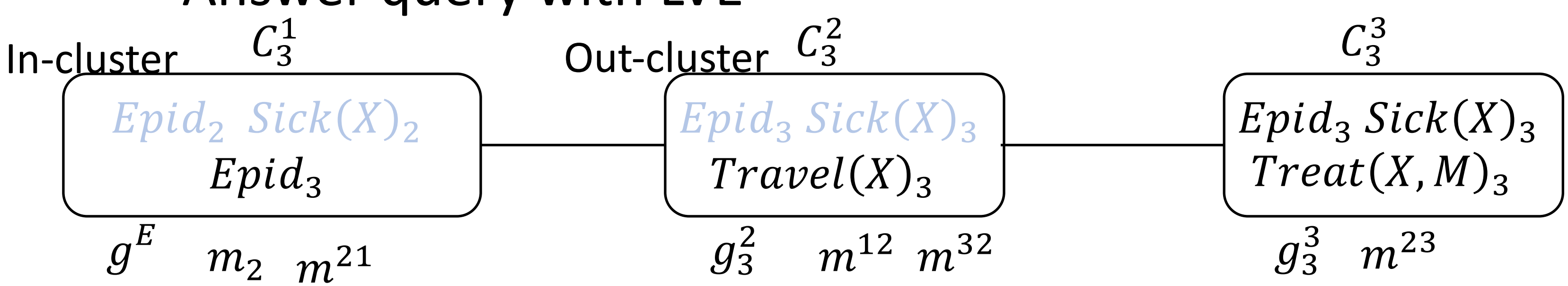
- Turn model in 1.5 time slice model
- Suffices to perform inference over time slice t
- From 1.5 time slice model construct FO jtree structure
- Ensure $\boldsymbol{I}_{t-1}$ is contained in a parcluster and $\boldsymbol{I}_t$ is contained in a parcluster
- Label parcluster with $\boldsymbol{I}_{t-1}$ as in-cluster and parcluster with $\boldsymbol{I}_t$ as out-



LDJT: Query answering

Gehrke et al. (2018)

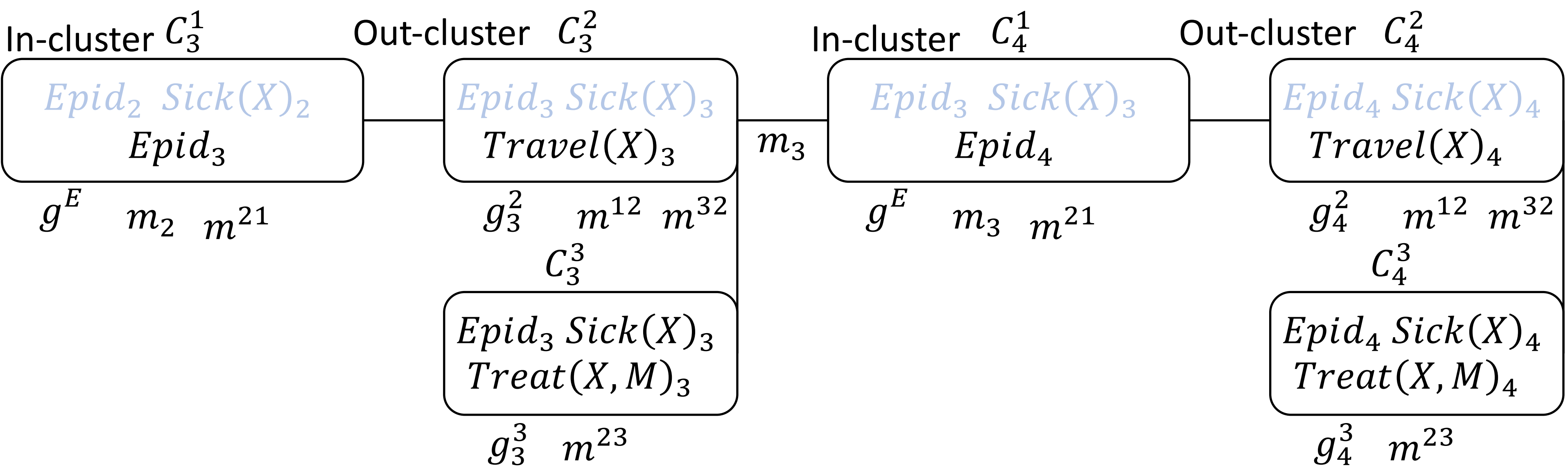
- Instantiate FO jtree structure
- Restore state description of interface variables
- Enter evidence
- Pass messages
- Query answering:
  - Find parcluster contain query term
  - Extract submodel
  - Answer query with LVE



LDJT: Proceed in time

Gehrke et al. (2018)

- Calculate m_3 using out-cluster (C_3^2)
- Eliminate $Travel(X)_3$ from C_3^2's local model
- Instantiate next FO jtree and enter m_3
- Enter evidence and pass messages



LDJT: Intermediate Overview

Gehrke et al. (2018)

- So far only a temporal forward pass
- Reason over one time step
- Keep only one time step in memory
- Filtering queries
- Prediction queries (filtering without new evidence)
- Hindsight queries

LDJT: Forward and Backward Pass

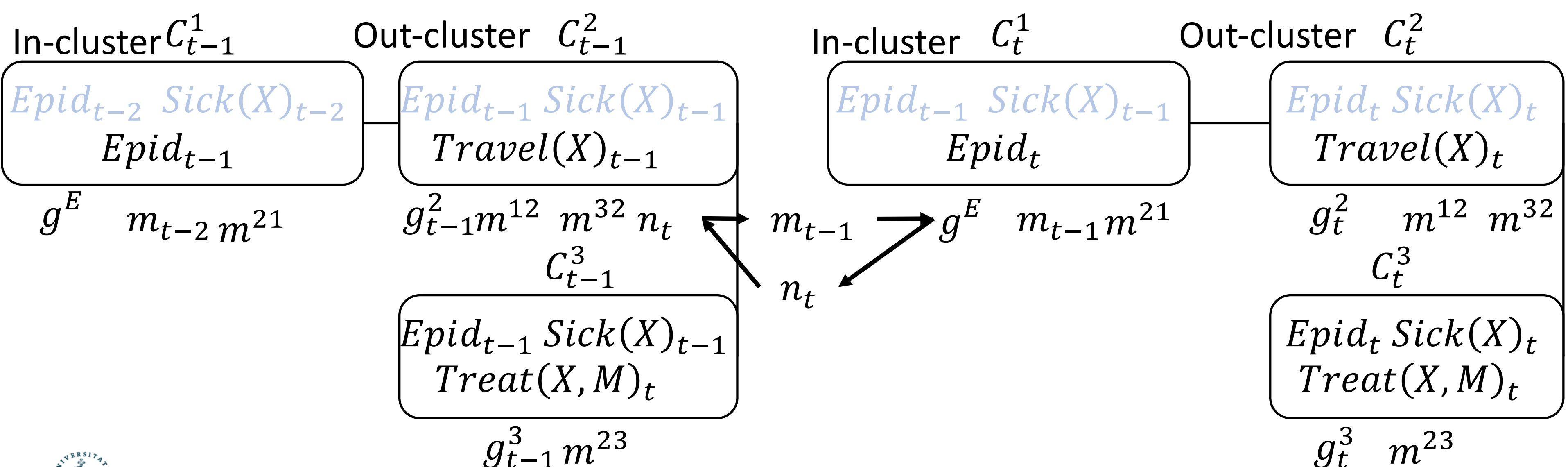
Gehrke et al. (2019)

- Use same FO jtree structures for backward pass
- Calculate a message n using an in-cluster over interface variables and pass n to previous time step
- LDJT needs to keep FO jtrees of previous time steps
- Different instantiation approaches during a backward pass
  - Keep all computations for all time steps in memory (not always feasible)
  - Instantiate time steps on demand (same as for the forward pass, possible due to the separation between time steps)

LDJT: Backward Pass

Gehrke et al. (2019)

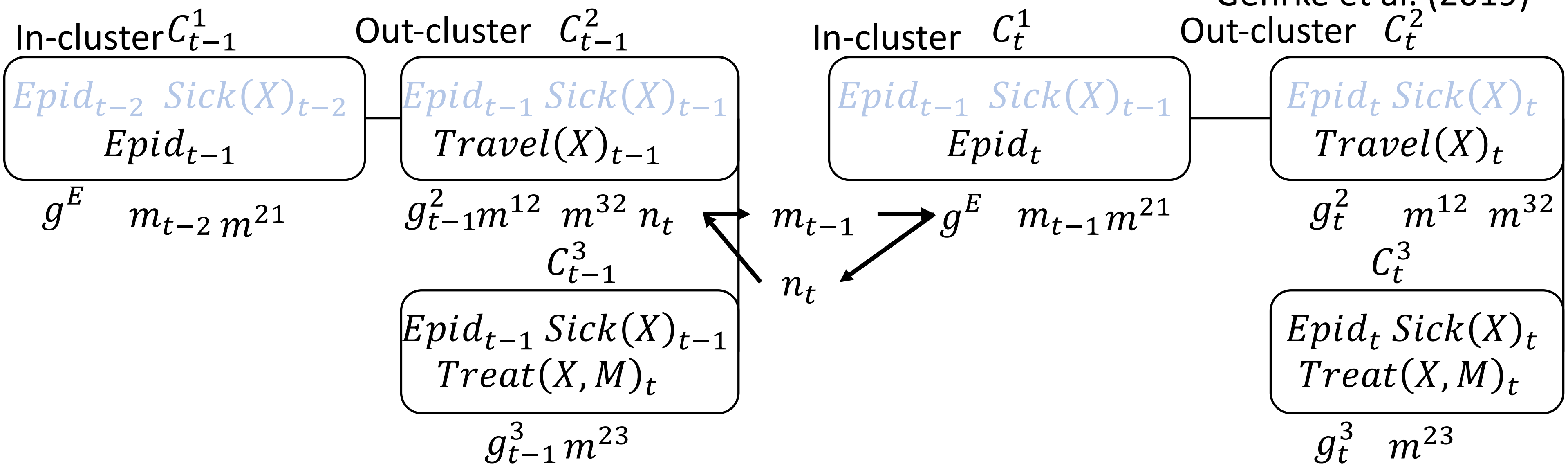
- Calculate n_t using in-cluster (C_t^1)
- Eliminate $Epid_t$ from C_t^1's local model, without m_{t-1}
- Add n_t to local model of out-cluster C_{t-1}^2
- Pass messages for $t-1$ to account for n_t



UNIVERSITÄT ZU LÜBECK

LDJT: Instantiations during a Backward Pass

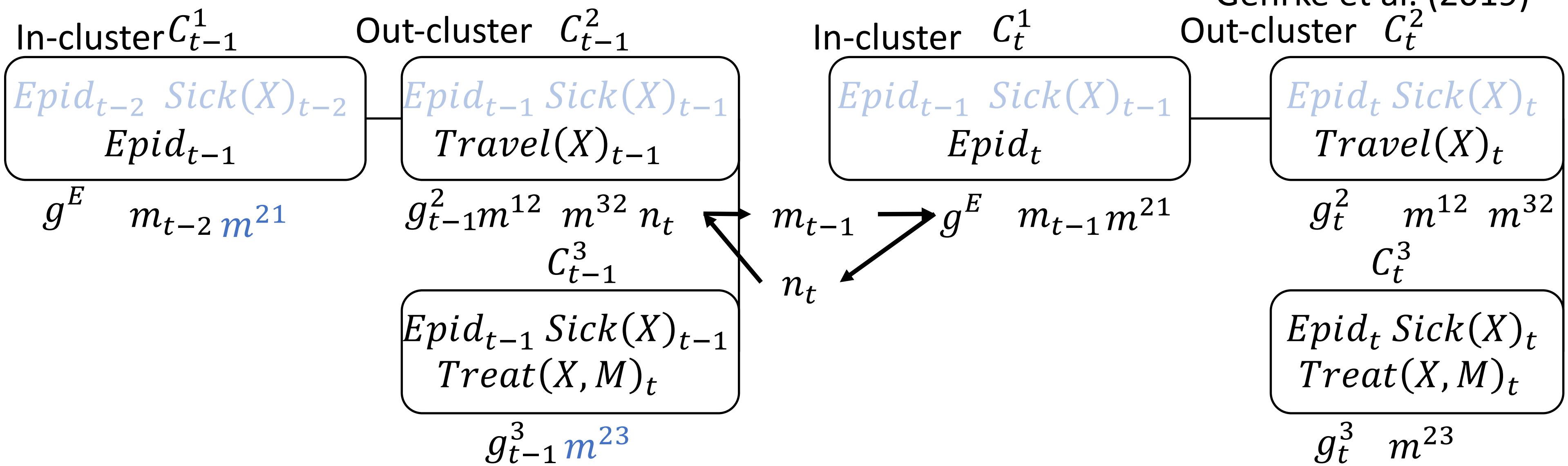
Gehrke et al. (2019)



	Keep Instantiations	Instantiate on demand
Messages to prepare for queries		
Messages to solely calculate n_{t-1}		
Additional memory for each time step		

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)

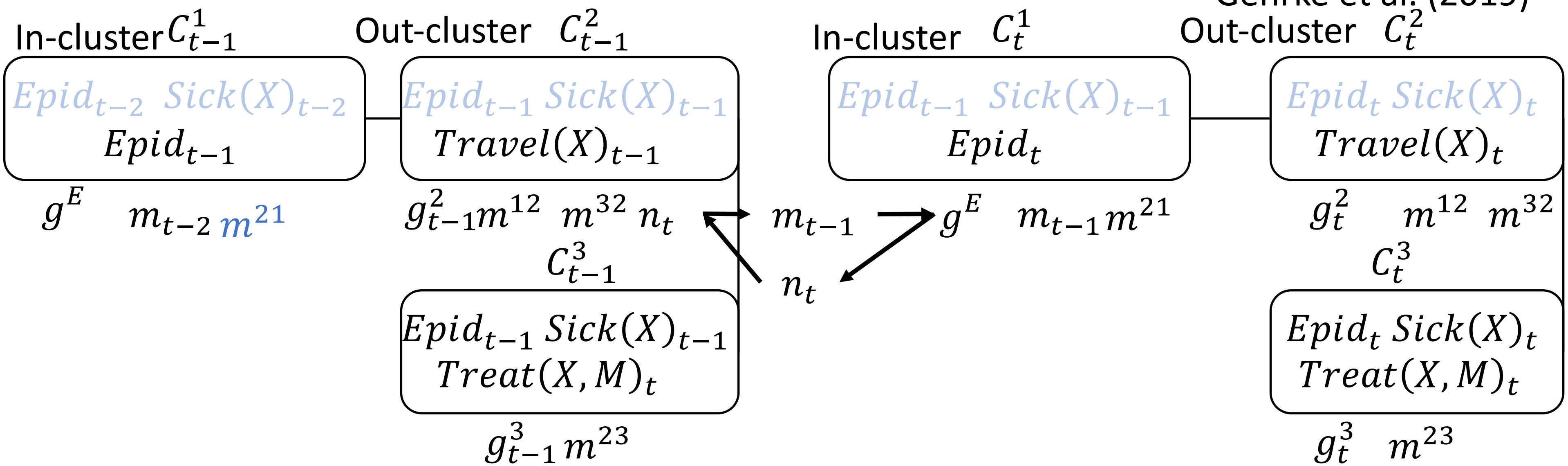


	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	
Messages to solely calculate n_{t-1}		
Additional memory for each time step		

n is the number of parclusters for each time step

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)

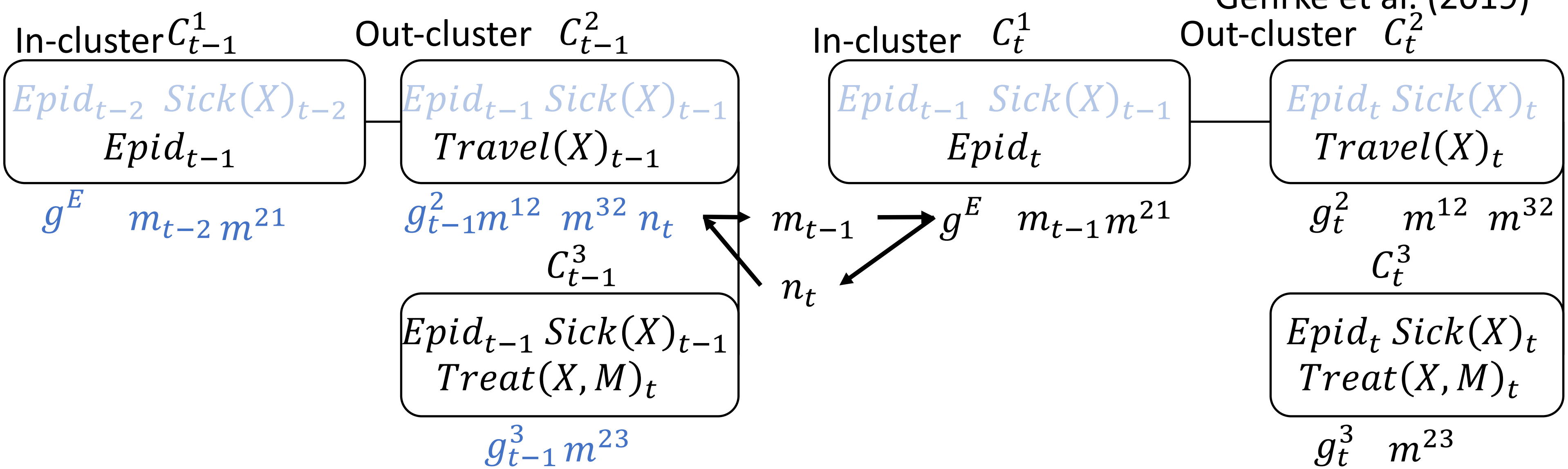


	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	
Messages to solely calculate n_{t-1}	$\leq n-1$	
Additional memory for each time step		

n is the number of parclusters for each time step

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)

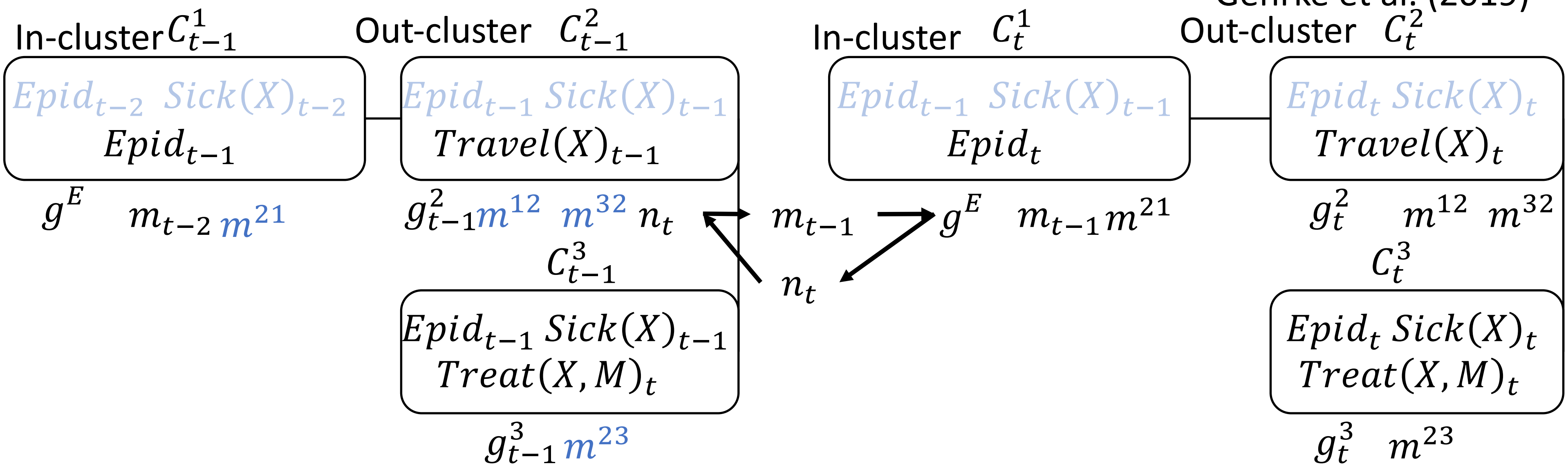


	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	
Messages to solely calculate n_{t-1}	$\leq n-1$	
Additional memory for each time step	All local models	

n is the number of parclusters for each time step

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)

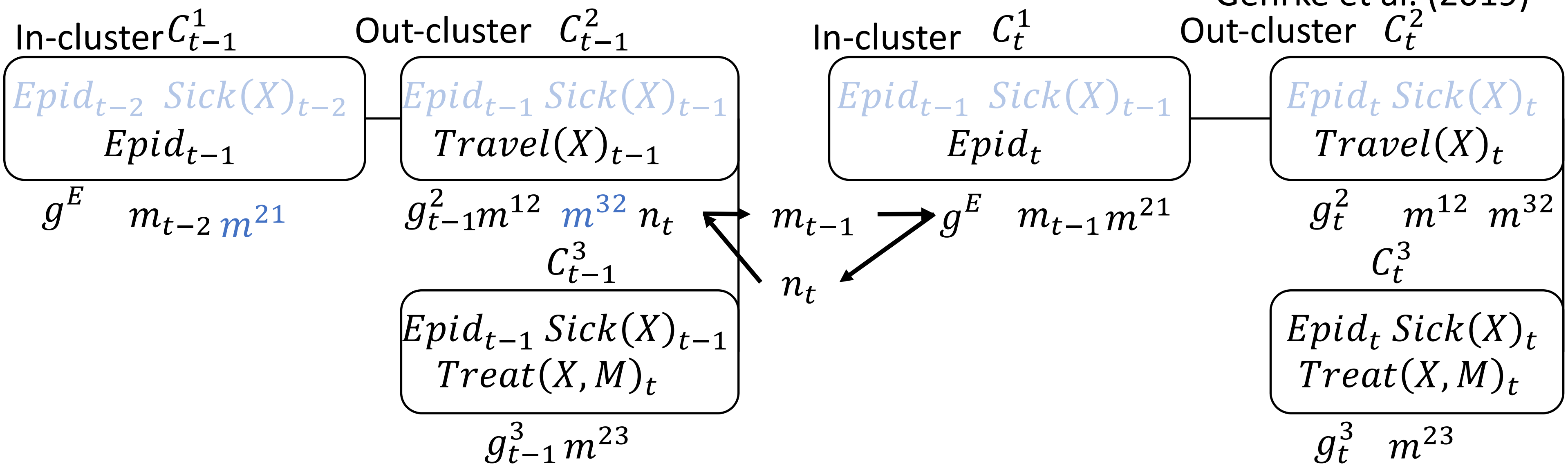


	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	$2*(n-1)$
Messages to solely calculate n_{t-1}	$\leq n-1$	
Additional memory for each time step	All local models	

n is the number of parclusters for each time step

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)

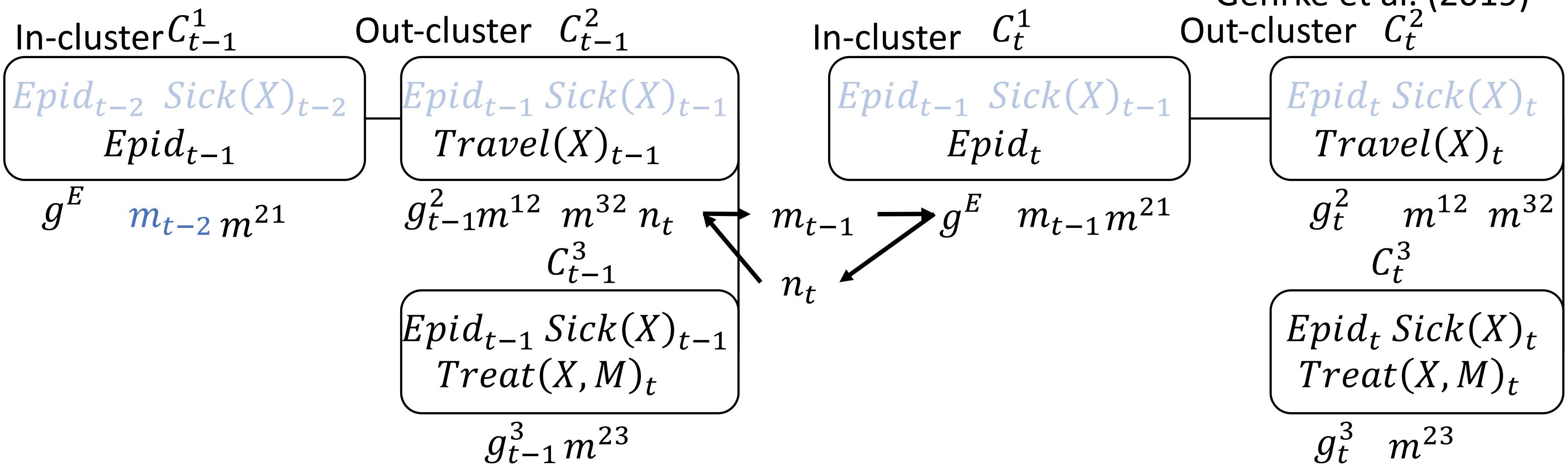


	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	$2*(n-1)$
Messages to solely calculate n_{t-1}	$\leq n-1$	$n-1$
Additional memory for each time step	All local models	

n is the number of parclusters for each time step

LDJT: Instantiations during a Backward Pass

Gehrke et al. (2019)



	Keep Instantiations	Instantiate on demand
Messages to prepare for queries	$n-1$	$2 \cdot (n-1)$
Messages to solely calculate n_{t-1}	$\leq n-1$	$n-1$
Additional memory for each time step	All local models	Only forward (m_t) messages

n is the number of parclusters for each time step

LDJT: Relational Forward Backward Algorithm

Gehrke et al. (2019)

- LDJT can answer hindsight queries, even to the first time step
- By combining the instantiation approaches, LDJT can trade off memory consumption and reusing computations
- LDJT is in the worst case quadratic to T, but normally remains linear w.r.t. T (T max # time steps)
- But does it really suffice to lift the interface algorithm?

LDJT: Preventing Unnecessary Groundings

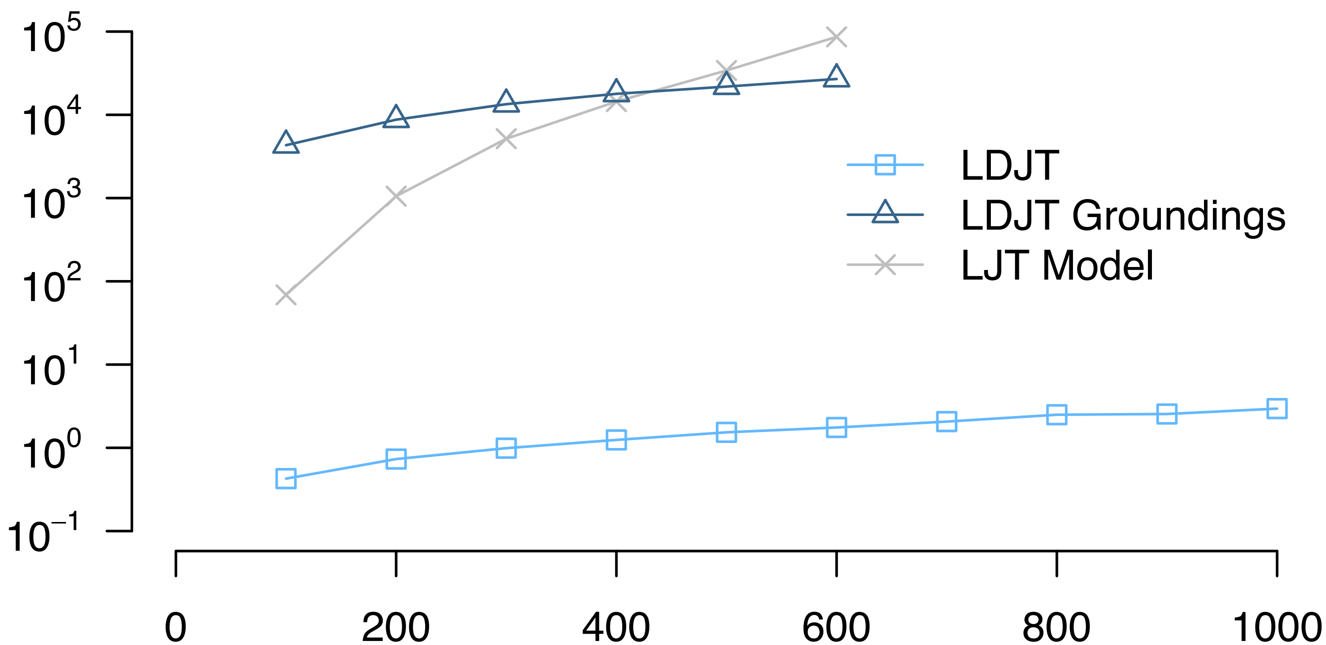
Gehrke et al. (2018b,c)

- Groundings in inter time slice messages (especially forward messages) can lead to grounding the model for all time steps
- Elimination order predetermined in FO jtree
- Non-ideal elimination order leads to groundings
  - Minimal set of interface variables not always ideal
  - Delay eliminations for inter time slice messages to prevent unnecessary groundings
  - Simply lifting the interface algorithm does not suffice, one also needs to ensure preconditions of lifting
- Trade off between lifting and handling temporal aspects due to restrictions on elimination orders

LDJT: Preventing Unnecessary Groundings

Gehrke et al. (2018b,c)

- Depending on the settings, either lifting or handling of temporal aspects is more efficient
- Preventing groundings to calculate a lifted solution pays off



LDJT: Theoretical Analysis

- FO^2 is not always liftable in temporal models
  - There exists an FO^2 for which LDJT has to ground
  - Unrolling would allow for a lifted solution
  - Handling temporal aspects restricts elimination order
- Lifting makes the problem manageable
  - Ground width grows with instances in interface
  - Lifted width remains the same
  - Runtime exponential to width

LDJT: Additional Queries

- Conjunctive queries over different time steps (Gehrke et al. (2018 d))
  - Can be used for event detection
  - What is the probability that someone travelled from X to Y and that afterwards there is a epidemic in Y given there is an epidemic in X?
- Maximum expected utility (Gehrke et al. (2019 b,c))
  - Decision support
  - Well studied within one time step (Apsel and Brafman (2011), Nath and Domingos (2009))
- Assignment queries (Gehrke et al. (2019 d))
  - Most likely state sequence
  - Well studied for static models (Dawid (1992), Dechter (1999), de Salvo Braz et al. (2006), Apsel and Brafman (2012), Braun and Möller (2018))

Taming Reasoning

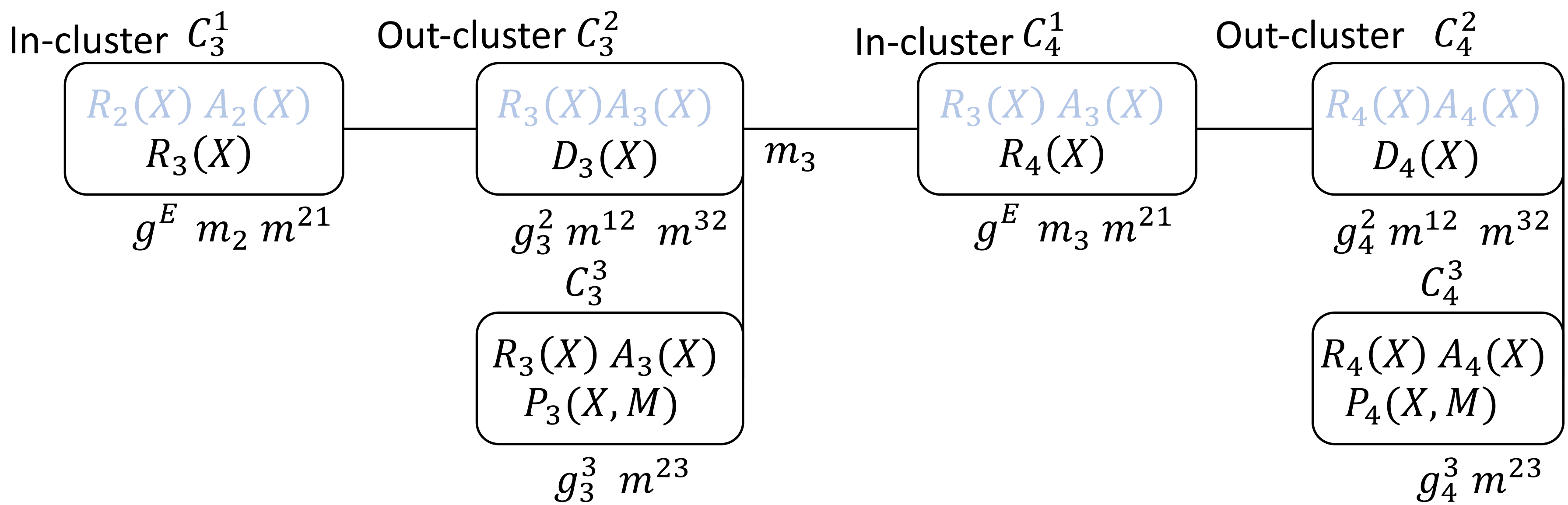
Gehrke et al. (2019e)

- Evidence can ground a model over time
- Non-symmetric evidence
  - Observe evidence for some instances in one time step
  - Observe evidence for a subset of these instances in another time step
  - Split the logical variable slowly over time
- Vanilla junction trees for each time step
- Forward message carries over splits, leading to slowly grounding a model over time

Evidence over Time

Gehrke et al. (2019e)

- $D_3(x_1) = true$
- Split g_3^2 into
  - $g_3^{2'}$ for x_1 and
  - $g_3^{2''}$ for $X \neq x_1$
- m_3 consists of
  - m^{12}
  - m^{32}
  - $g_3^{2'}$ and $g_3^{2''}$ with $D_3(X)$ eliminated



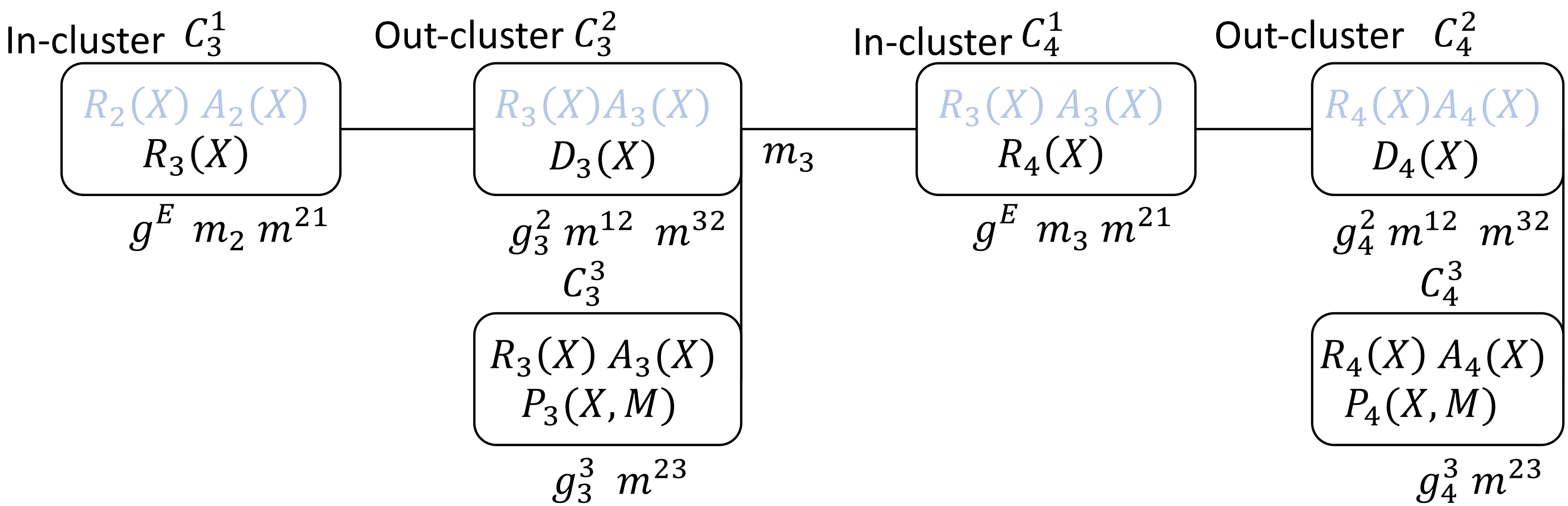
UNIVERSITÄT ZU LÜBECK

Evidence over Time

Gehrke et al. (2019e)

- $D_4(x_2) = true$
- Split g_4^2 into
  - $g_4^{2'}$ for x_2 and
  - $g_4^{2''}$ for $X \neq x_2$

- m_4 consists of
  - m^{12} (containing m_3)
  - m^{32}
  - $g_4^{2'}$ and $g_4^{2''}$ with $D_4(X)$ eliminated



Undoing Splits

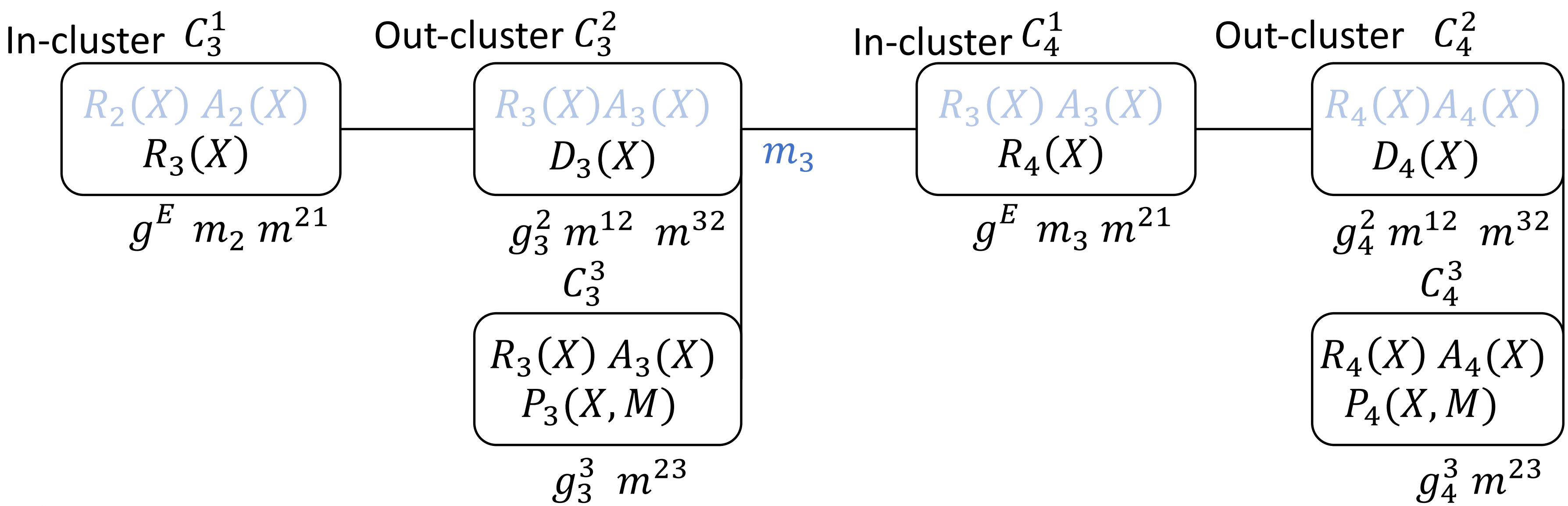
Gehrke et al. (2019e)

- Need to undo splits to keep reasoning polynomial w.r.t. domain sizes
- Where can splits be undone efficiently?
- How to undo splits?
- Is it reasonable to undo splits?

Where Can Splits Be Undone Efficiently?

Gehrke et al. (2019e)

- Evidence causes splits in a logical variable in the same way in all factors in a model
- LDJT always instantiates a vanilla junction tree
- Forward message carries over splits



How to Undo Splits?

Gehrke et al. (2019e)

- The colouring algorithm (Ahmadi et al. 2013) can efficiently identify exact symmetries
- Evidence causes differences in distributions
- Need to find approximate symmetries to undo splits caused by evidence
- Need a way to merge factors

Comparing Factors

Gehrke et al. (2019e)

- Comparing all marginals is expensive
- Comparing marginals of a subset of random variables can determine non-similar factors similar

$R(X)$	$A(X)$	f
false	*false*	0
false	*true*	7
true	*false*	4
true	*true*	1

$R(X)$	$A(X)$	f
false	*false*	2
false	*true*	4
true	*false*	2
true	*true*	4

- $P(A(x_1 = true))$: $\frac{2}{3}$ $\frac{2}{3}$
- $P(R(x_1 = true))$: $\frac{5}{12}$ $\frac{1}{2}$

$|\mathcal{D}(X)|$=1

Comparing Factors

Gehrke et al. (2019e)

- Potentials determine distributions
- Similar ratios in potentials lead to similar marginals and similar factors

$R(X)$	$A(X)$	f
false	*false*	4
false	*true*	3
true	*false*	2
true	*true*	1

$R(X)$	$A(X)$	f
false	*false*	3.9
false	*true*	3.1
true	*false*	2.1
true	*true*	0.9

- $P(A(x_1 = true))$: $\frac{4}{10}$ $\frac{4}{10}$
- $P(R(x_1 = true))$: $\frac{3}{10}$ $\frac{3}{10}$
- $P(A(x_1 = true) \wedge R(x_1 = true)$: $\frac{1}{10}$ $\frac{0.9}{10}$

$|\mathcal{D}(X)|$=1

Find Approximate Symmetries

Gehrke et al. (2019e)

- Cosine similarity for similarity between vector
- $\cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$

$R(X)$	$A(X)$	f
false	*false*	0
false	*true*	7
true	*false*	4
true	*true*	1

$R(X)$	$A(X)$	f
false	*false*	2
false	*true*	4
true	*false*	2
true	*true*	4

- $\cos(\theta) = \frac{0 \cdot 2 + 7 \cdot 4 + 4 \cdot 2 + 1 \cdot 4}{\sqrt{0+49+16+1} \cdot \sqrt{4+16+4+16}} \sim 0.7785$

UNIVERSITÄT ZU LÜBECK

Find Approximate Symmetries

Gehrke et al. (2019e)

- Cosine similarity for similarity between vector
- $\cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$

$R(X)$	$A(X)$	f
false	*false*	4
false	*true*	3
true	*false*	2
true	*true*	1

$R(X)$	$A(X)$	f
false	*false*	3.9
false	*true*	3.1
true	*false*	2.1
true	*true*	0.9

- $\cos(\theta) = \frac{4 \cdot 3.9 + 3 \cdot 3.1 + 2 \cdot 2.1 + 1 \cdot 0.9}{\sqrt{16+9+4+1} \cdot \sqrt{15.21+9.61+4.41+0.81}} \sim 0.9993$

Find Approximate Symmetries

Gehrke et al. (2019e)

- Cosine similarity for similarity between vector
- $\cos(\theta) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$

$R(X)$	$A(X)$	f
false	*false*	4
false	*true*	3
true	*false*	2
true	*true*	1

$R(X)$	$A(X)$	f
false	*false*	8
false	*true*	6
true	*false*	4
true	*true*	2

- $\cos(\theta) = \frac{4 \cdot 8 + 3 \cdot 6 + 2 \cdot 4 + 1 \cdot 3}{\sqrt{16+9+4+1} \cdot \sqrt{64+36+16+4}} = 1$
- Cluster splits with 1-cos as distance function

UNIVERSITÄT ZU LÜBECK

Merging Clusters

Gehrke et al. (2019e)

- Merge identified clusters based on distance function while accounting for groundings

$|\mathcal{D}(X)|$= 4

$R(X)$	$A(X)$	f
false	*false*	4
false	*true*	3
true	*false*	2
true	*true*	1

$|\mathcal{D}(X')|$= 4

$R(X')$	$A(X')$	f
false	*false*	7.9
false	*true*	6
true	*false*	3.9
true	*true*	2.1

$|\mathcal{D}(X'')|$= 2

$R(X'')$	$A(X'')$	f
false	*false*	15.7
false	*true*	12.2
true	*false*	8.1
true	*true*	3.8

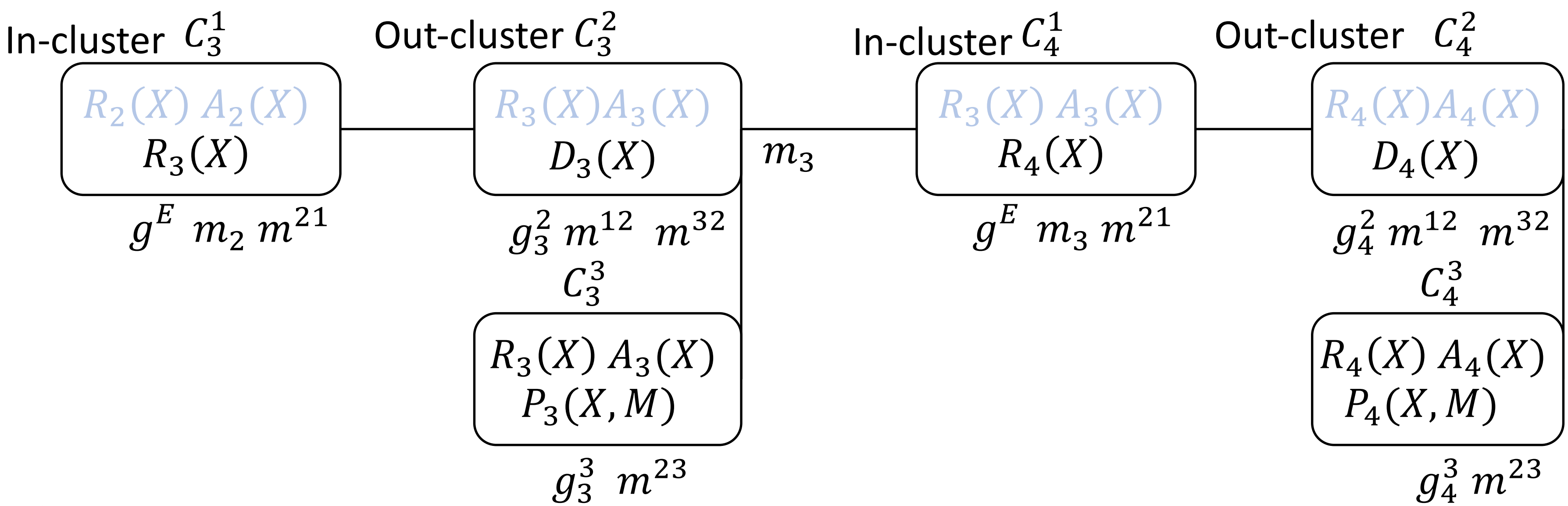
$|\mathcal{D}(X)|$= 10

$R(X)$	$A(X)$	f
false	*false*	$\frac{(4\cdot4+7.9\cdot4+15.7\cdot2)}{10} = 7.9$
false	*true*	$\frac{(3\cdot4+6\cdot4+12.2\cdot2)}{10} = 6.04$
true	*false*	$\frac{(2\cdot4+3.9\cdot4+8.1\cdot2)}{10} = 3.98$
true	*true*	$\frac{(1\cdot4+2.1\cdot4+3.8\cdot2)}{10} = 2$

Is It Reasonable to Undo Splits?

Gehrke et al. (2019e)

- Approximate forward message
- For each time step the temporal behaviour is multiplied on the forward message
- Indefinitely bounded error due to temporal behaviour



Taming Reasoning

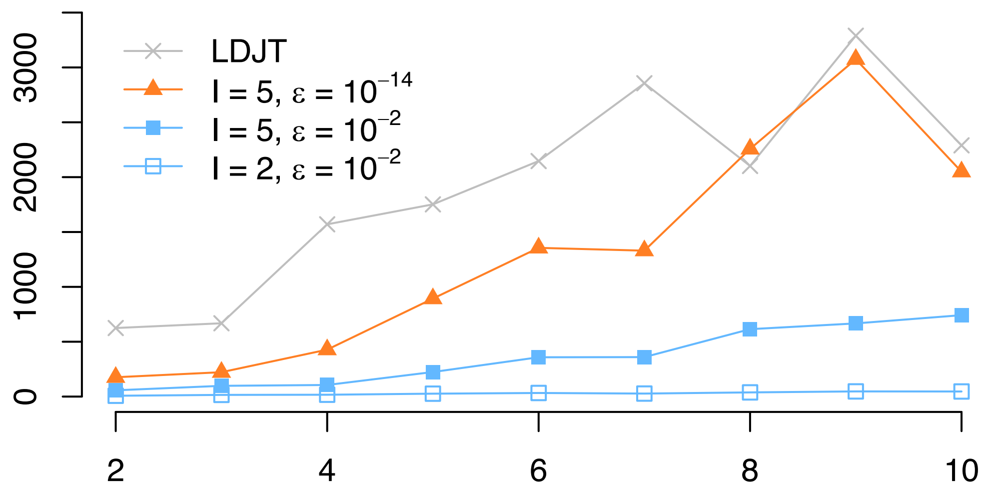
Gehrke et al. (2019e)

- Need to undo splits to keep reasoning polynomial w.r.t. domain sizes
- Where can splits be undone efficiently?
  - Undo splits in a forward message
- How to undo splits?
  - Find approximate symmetries
  - Merge based on groundings
- Is it reasonable to undo splits
  - Yes, due to the temporal model behaviour (indefinitely bounded error)

Results

Gehrke et al. (2019e)

- DBSCAN for Clustering
- ANOVA for checking fitness of clusters



π	Max	Min	Average
0	*0.0001537746121*	0.0000000001720	0.0000191206488
2	*0.0000000851654*	0.0000000000001	0.0000000111949
4	*0.0000000000478*	0	0.0000000000068

Outlook

- Continue optimising
  - Parallelisation
  - Caching
- From discrete time interval to time continuous
- Preserving symmetries
- Learning?
  - Structure
  - Potentials (Idea of Baum Welch now possible)
  - Symmetries
  - *Transfer learning*
- Open world?
  - Unknown domains
  - Unknown behaviour

Wrap-up Exact Lifted Dynamic Inference

- Parfactor models for sparse encoding
  - Factorisation of full joint distribution
  - Logical variables to model objects
- Algorithms for exact query answering
  - LDJT for repeated inference
  - Extensions possible
    - Parameterised, conjunctive queries
    - Maximum expected utility
    - Assignment queries

Next: Summary

References

- 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.*
- Apsel and Brafman (2012)
  Udi Apsel and Ronen I. Brafman. Exploiting Uniform Assignments in First-Order MPE. *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.
- Apsel and Brafman (2011)
  Udi Apsel and Ronen I. Brafman. Extended Lifted Inference with Joint Formulas. Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. pp. 11–18, 2011.
- Dawid (1992)
  Alexander Philip Dawid. Applications of a General Propagation Algorithm for Probabilistic Expert Systems. *Statistics and Computing*, 2(1):25–36, 1992.

References

- Dechter (1999)
  Rina Dechter. Bucket Elimination: A Unifying Framework for Probabilistic Inference. In *Learning and Inference in Graphical Models*, pages 75–104. MIT Press, 1999.
- De Salvo Braz et al. (2006)
  Rodrigo de Salvo Braz, Eyal Amir, and Dan Roth. MPE and Partial Inversion in Lifted Probabilistic Variable Elimination. *AAAI-06 Proceedings of the 21st Conference on Artificial Intelligence*, 2006.
- Murphy (2002)
  Kevin P. Murphy. Dynamic Bayesian Networks: Representation, Inference and Learning. *PhD Thesis University of California, Berkeley*, 2002.
- Nath and Domingos (2009)
  Aniruddh Nath and Pedro Domingos, A language for relational decision theory, Proceedings of the International Workshop on Statistical Relational Learning, 2009.

Work @ IFIS

- Braun and Möller (2018b)
  Tanya Braun and Ralf Möller. Lifted Most Probable Explanation. In *Proceedings of the International Conference on Conceptual Structures*, 2018.
- Gehrke et al. (2018)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Lifted Dynamic Junction Tree Algorithm. In *Proceedings of the International Conference on Conceptual Structures*, 2018.
- Gehrke et al. (2018b)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Towards Preventing Unnecessary Groundings in the Lifted Dynamic Junction Tree Algorithm. In Proceedings of KI 2018: Advances in Artificial Intelligence, 2018.
- Gehrke et al. (2018c)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Preventing Unnecessary Groundings in the Lifted Dynamic Junction Tree Algorithm. In Proceedings of the AI 2018: Advances in Artificial Intelligence, 2018.

Work @ IFIS

- 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.
- Gehrke et al. (2019b)
  Marcel Gehrke, Tanya Braun, Ralf Möller, Alexander Waschkau, Christoph Strumann, and Jost Steinhäuser. Lifted Maximum Expected Utility. In Artificial Intelligence in Health, 2019.
- Gehrke et al. (2019c)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Lifted Temporal Maximum Expected Utility. In Proceedings of the 32nd Canadian Conference on Artificial Intelligence, Canadian AI 2019, 2019.

Work @ IFIS

- Gehrke et al. (2019d)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Lifted Temporal Most Probable Explanation In Proceedings *of the International Conference on Conceptual Structures*, 2019
- Gehrke et al. (2019e)
  Marcel Gehrke, Tanya Braun, and Ralf Möller. Lifted Taming Reasoning in Temporal Probabilistic Relational Models *Technical report*