On Decidability and Tractability of Querying in Temporal EL

Roman Kontchakov

Department of Computer Science and Inf. Systems, Birkbeck College, London

http://www.dcs.bbk.ac.uk/~roman

joint work with **Víctor Gutiérrez-Basulto** and **Jean Christoph Jung**

(*Universität Bremen*)

OBDA with Temporal Data

query $q(x) = \mathsf{RequiresBloodTest}(x, \mathsf{today})$

ontology

$\mathsf{Patient} \sqcap \bigcirc_P^5 \exists \mathsf{vaccinated}.\mathsf{LiveVirus} \sqsubseteq \mathsf{ViableParticipant}$

$\mathsf{Patient} \sqcap \mathsf{RequiresBloodTest} \sqsubseteq \bigcirc_F^3 \mathsf{RequiresBloodTest}$

ABox

Patient(*john*, 21/04/16)
RequiresBloodTest(*john*, 18/04/16)
vaccinated(*john*, *measles*, 16/04/16)
LiveVirus(*measles*, 01/01/16)
. . .

mappings

timestamped data: vaccinations, blood tests, etc.

Temporal $\mathcal{EL}$

concepts $C, D \; ::= \; A \mid C \sqcap D \mid \exists r.C \mid \underbrace{\bigcirc_F C}_{\text{at the next moment}} \mid \underbrace{\bigcirc_F C}_{\text{at the previous moment}}$

$\mid \underbrace{\Diamond_F C}_{\text{sometime in the future}} \mid \underbrace{\Diamond_P C}_{\text{sometime in the past}}$

roles r are partitioned into **local** and **rigid** (time-invariant)

Temporal $\mathcal{EL}$

concepts $C, D \ ::= \ A \ \mid \ C \sqcap D \ \mid \ \exists r.C \ \mid \ \underbrace{\bigcirc_F C}_{\text{at the next moment}} \ \mid \ \underbrace{\bigcirc_F C}_{\text{at the previous moment}}$

$\mid \ \underbrace{\Diamond_F C}_{\text{sometime in the future}} \ \mid \ \underbrace{\Diamond_P C}_{\text{sometime in the past}}$

roles r are partitioned into **local** and **rigid** (time-invariant)

concept inclusions: $C \sqcap \bigcirc_F C \sqcap \bigcirc_F \bigcirc_F C \sqsubseteq B$ and $\Diamond_P D \sqsubseteq U$

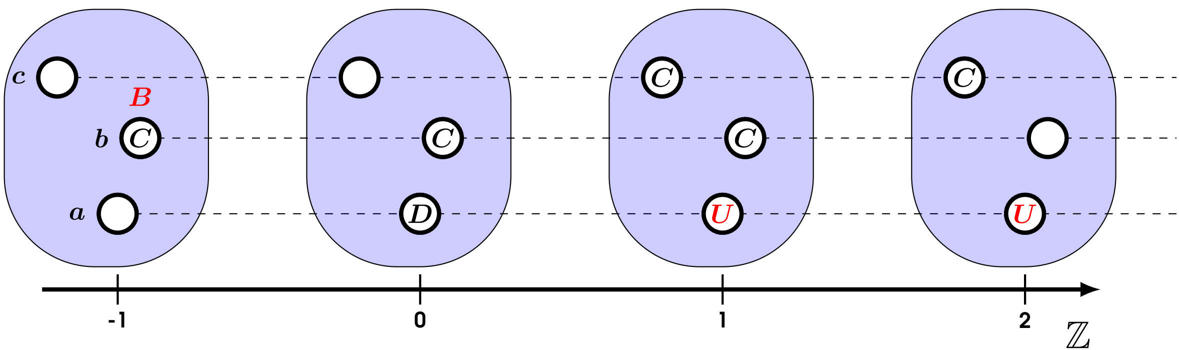
Temporal $\mathcal{EL}$

concepts $C, D \;::=\; A \mid C \sqcap D \mid \exists r.C \mid \underbrace{\bigcirc_F C}_{\text{at the next moment}} \mid \underbrace{\bigcirc_F C}_{\text{at the previous moment}}$

$\mid \underbrace{\Diamond_F C}_{\text{sometime in the future}} \mid \underbrace{\Diamond_P C}_{\text{sometime in the past}}$

roles r are partitioned into **local** and **rigid** (time-invariant)

concept inclusions: $C \sqcap \bigcirc_F C \sqcap \bigcirc_F \bigcirc_F C \sqsubseteq B$ and $\Diamond_P D \sqsubseteq U$



- constant domains with rigid interpretation of individuals

  under the standard name assumption
- concept inclusions hold at all moments of time (globally)

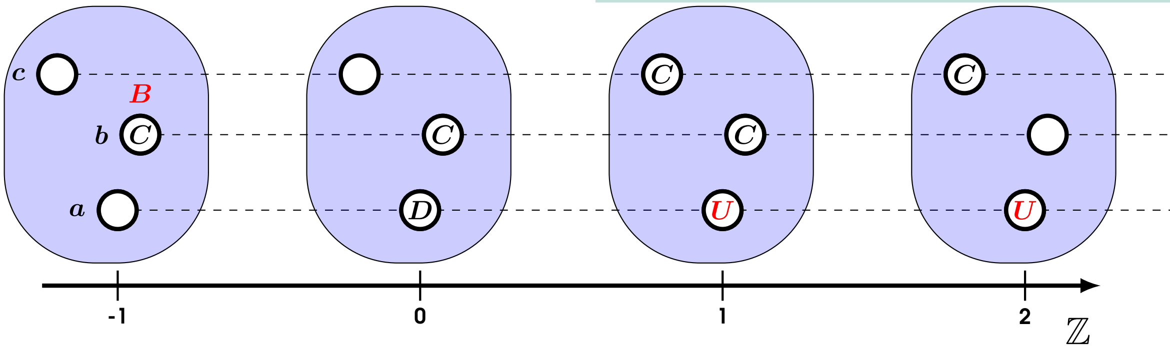
Temporal $\mathcal{EL}$

concepts $C, D \;::=\; A \;\mid\; C \sqcap D \;\mid\; \exists r.C \;\mid\; \underbrace{\bigcirc_F C}_{\text{at the next moment}} \;\mid\; \underbrace{\bigcirc_F C}_{\text{at the previous moment}} \;\mid\; \underbrace{\Diamond_F C}_{\text{sometime in the future}} \;\mid\; \underbrace{\Diamond_P C}_{\text{sometime in the past}}$

roles r are partitioned into **local** and **rigid** (time-invariant)

concept inclusions: $C \sqcap \bigcirc_F C \sqcap \bigcirc_F \bigcirc_F C \sqsubseteq B$ and $\Diamond_P D \sqsubseteq U$

$D \sqsubseteq D'$, $D' \sqsubseteq \bigcirc_F U$, $\bigcirc_P D' \sqsubseteq D'$



- constant domains with rigid interpretation of individuals
  under the standard name assumption
- concept inclusions hold at all moments of time (globally)

Querying Temporal $\mathcal{EL}$

temporal atomic query $A(x,t)$ $\quad$ A is a concept name

entailment for $a \in \mathsf{ind}(\mathcal{A})$ and $i \in \mathbb{Z}$ $\quad$ time instants are represented in unary

$$\mathcal{T},\mathcal{A} \models A(a,i) \quad \text{iff} \quad \mathfrak{I} \models A(a,i), \quad \text{for all models } \mathfrak{I} \text{ of } \mathcal{T},\mathcal{A}$$

certain answers $\quad (a,n) \in \mathsf{ind}(\mathcal{A}) \times \underbrace{\mathsf{tem}(\mathcal{A})}_{\min \mathcal{A} \leq n \leq \max \mathcal{A}}$ such that $\mathcal{T},\mathcal{A} \models A(a,n)$

Querying Temporal $\mathcal{EL}$

temporal atomic query $A(x, t)$ A is a concept name

entailment for $a \in \mathsf{ind}(\mathcal{A})$ and $i \in \mathbb{Z}$ time instants are represented in unary

$$\mathcal{T}, \mathcal{A} \models A(a, i) \quad \text{iff} \quad \mathfrak{J} \models A(a, i), \quad \text{for all models } \mathfrak{J} \text{ of } \mathcal{T}, \mathcal{A}$$

certain answers $(a, n) \in \mathsf{ind}(\mathcal{A}) \times \underbrace{\mathsf{tem}(\mathcal{A})}_{\min \mathcal{A} \leq n \leq \max \mathcal{A}}$ such that $\mathcal{T}, \mathcal{A} \models A(a, n)$

Bad News

Theorem TAQ answering over full $\mathcal{TEL}$ is undecidable

(even for **data** complexity, with a fixed TBox and a fixed query)

hint: express $\sqcup$ using $\Diamond_F$ on the right-hand side of CIs

Querying Temporal $\mathcal{EL}$

temporal atomic query $A(x, t)$ A is a concept name

entailment for $a \in \mathsf{ind}(\mathcal{A})$ and $i \in \mathbb{Z}$ time instants are represented in unary

$$\mathcal{T}, \mathcal{A} \models A(a, i) \quad \text{iff} \quad \mathfrak{J} \models A(a, i), \quad \text{for all models } \mathfrak{J} \text{ of } \mathcal{T}, \mathcal{A}$$

certain answers $(a, n) \in \mathsf{ind}(\mathcal{A}) \times \underbrace{\mathsf{tem}(\mathcal{A})}_{\min \mathcal{A} \leq n \leq \max \mathcal{A}}$ such that $\mathcal{T}, \mathcal{A} \models A(a, n)$

Bad News

Theorem TAQ answering over full $\mathcal{TEL}$ is undecidable
(even for **data** complexity, with a fixed TBox and a fixed query)

hint: express $\sqcup$ using $\Diamond_F$ on the right-hand side of CIs

$\mathcal{TEL}^{\bigcirc}$ is $\mathcal{TEL}$ without $\Diamond_*$

Theorem TAQ answering over $\mathcal{TEL}^{\bigcirc}$ with functional roles
is undecidable for data complexity

Theorem TAQ answering over $\mathcal{TEL}^{\bigcirc}$ with inverse roles
is undecidable for data complexity

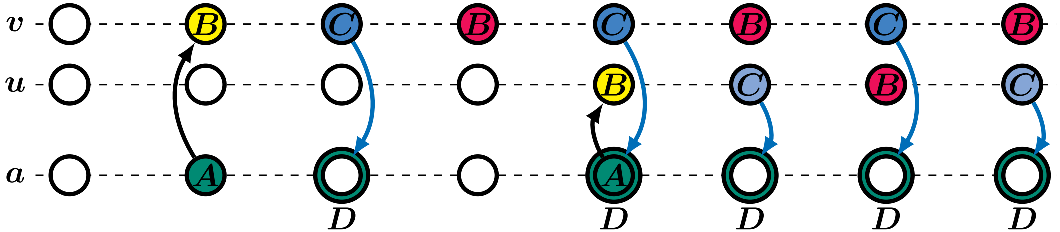
(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B, \qquad \bigcirc_{*} B \sqsubseteq A, \qquad A \sqsubseteq \exists r.B, \qquad \exists r.B \sqsubseteq A$

(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B$, $\quad \bigcirc_* B \sqsubseteq A$, $\quad A \sqsubseteq \exists r.B$, $\quad \exists r.B \sqsubseteq A$

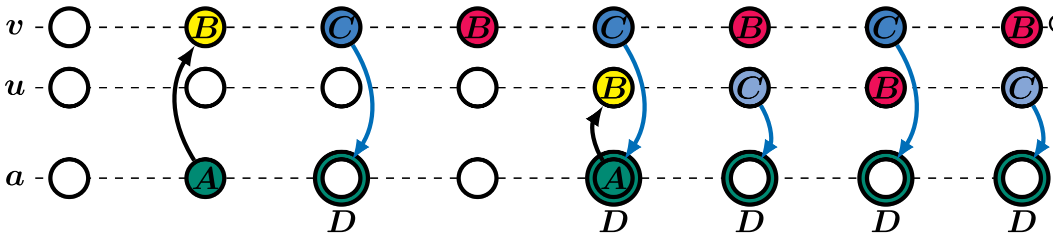
Example: $\bigcirc_P B \sqsubseteq C$, $\bigcirc_P C \sqsubseteq B$, $A \sqsubseteq \exists r.B$, $\exists r.C \sqsubseteq D$ with rigid r



(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B$, $\bigcirc_* B \sqsubseteq A$, $A \sqsubseteq \exists r.B$, $\exists r.B \sqsubseteq A$

Example: $\bigcirc_P B \sqsubseteq C$, $\bigcirc_P C \sqsubseteq B$, $A \sqsubseteq \exists r.B$, $\exists r.C \sqsubseteq D$ with rigid r



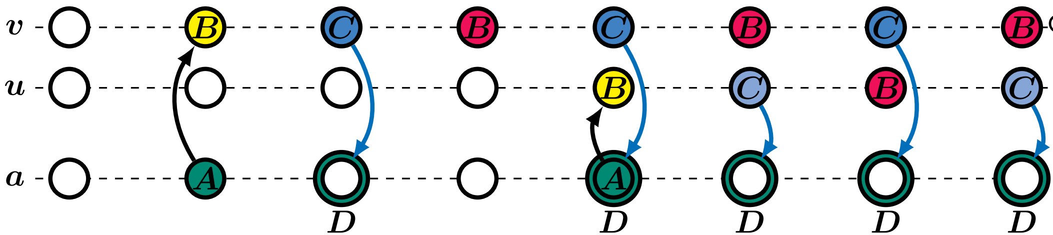
two copies of the same trace with different **shifts**

a trace is a map $\pi : \mathbb{Z} \to \mathsf{CN}$ that respects all $A \sqcap A' \sqsubseteq B$ and $\bigcirc_* B \sqsubseteq A$

(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B$, $\quad \bigcirc_{*} B \sqsubseteq A$, $\quad A \sqsubseteq \exists r.B$, $\quad \exists r.B \sqsubseteq A$

Example: $\bigcirc_{P} B \sqsubseteq C$, $\bigcirc_{P} C \sqsubseteq B$, $A \sqsubseteq \exists r.B$, $\exists r.C \sqsubseteq D$ with rigid r



two copies of the same trace with different shifts

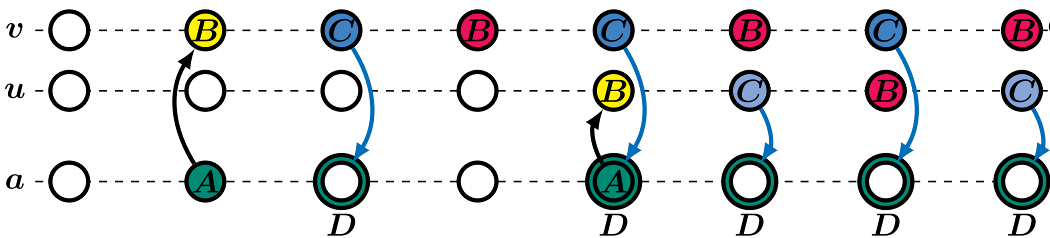
a trace is a map $\pi : \mathbb{Z} \to \mathsf{CN}$ that respects all $A \sqcap A' \sqsubseteq B$ and $\bigcirc_{*} B \sqsubseteq A$

quasimodel = traces π_a for $a \in \mathsf{ind}(\mathcal{A})$ + traces π_B for $B \in \mathsf{CN}$
that contains the ABox $\mathcal{A}$ in the π_a and $B \in \pi_B(0)$
and respects all $\exists r.B \sqsubseteq A$

(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B$, $\quad \bigcirc_* B \sqsubseteq A$, $\quad A \sqsubseteq \exists r.B$, $\quad \exists r.B \sqsubseteq A$

Example: $\bigcirc_P B \sqsubseteq C$, $\bigcirc_P C \sqsubseteq B$, $A \sqsubseteq \exists r.B$, $\exists r.C \sqsubseteq D$ with rigid r



two copies of the same trace with different shifts

a trace is a map $\pi : \mathbb{Z} \to \mathsf{CN}$ that respects all $A \sqcap A' \sqsubseteq B$ and $\bigcirc_* B \sqsubseteq A$

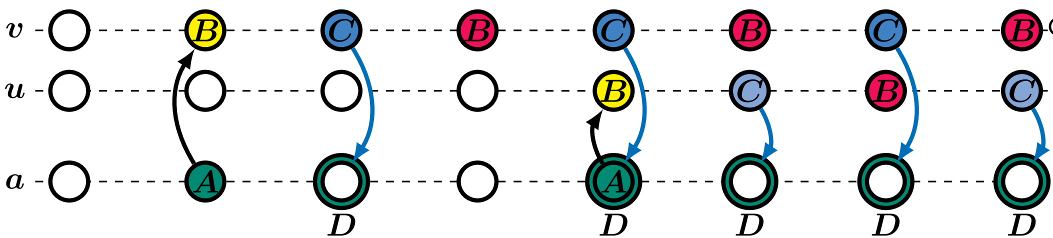
quasimodel = traces π_a for $a \in \mathsf{ind}(\mathcal{A})$ + traces π_B for $B \in \mathsf{CN}$
that contains the ABox $\mathcal{A}$ in the π_a and $B \in \pi_B(0)$
and respects all $\exists r.B \sqsubseteq A$

Theorem $\mathcal{T}, \mathcal{A} \models A(a, i)$ iff $A \in \pi_a(i)$, for $a \in \mathsf{ind}(\mathcal{A})$ and $i \in \mathbb{Z}$

(Some) Good News

normal form: $A \sqcap A' \sqsubseteq B$, $\quad \bigcirc_* B \sqsubseteq A$, $\quad A \sqsubseteq \exists r.B$, $\quad \exists r.B \sqsubseteq A$

Example: $\bigcirc_P B \sqsubseteq C$, $\bigcirc_P C \sqsubseteq B$, $A \sqsubseteq \exists r.B$, $\exists r.C \sqsubseteq D$ with rigid r



two copies of the same trace with different shifts

a trace is a map $\pi : \mathbb{Z} \to \mathsf{CN}$ that respects all $A \sqcap A' \sqsubseteq B$ and $\bigcirc_* B \sqsubseteq A$

quasimodel = traces π_a for $a \in \mathsf{ind}(\mathcal{A})$ + traces π_B for $B \in \mathsf{CN}$

that contains the ABox $\mathcal{A}$ in the π_a and $B \in \pi_B(0)$

and respects all $\exists r.B \sqsubseteq A$

Theorem $\mathcal{T}, \mathcal{A} \models A(a, i)$ iff $A \in \pi_a(i)$, for $a \in \mathsf{ind}(\mathcal{A})$ and $i \in \mathbb{Z}$

traces are infinite structures

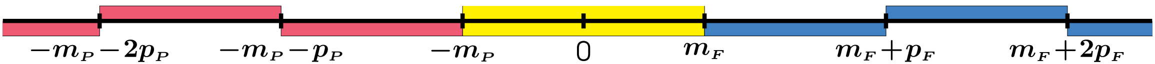
Ultimate Periodicity

a quasimodel is **ultimately p-periodic** if

for each trace π, there are positive integers $m_P, p_P, m_F, p_F \leq p$ such that

$\pi_B(n - p_P) = \pi_B(n)$, for all $n \leq -m_P$

$\pi_B(n + p_F) = \pi_B(n)$, for all $n \geq m_F$



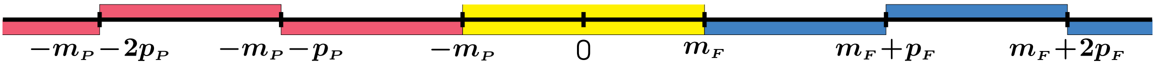
Ultimate Periodicity

a quasimodel is **ultimately p-periodic** if

for each trace π, there are positive integers $m_P, p_P, m_F, p_F \leq p$ such that

$\pi_B(n - p_P) = \pi_B(n)$, for all $n \leq -m_P$

$\pi_B(n + p_F) = \pi_B(n)$, for all $n \geq m_F$



Theorem TAQ answering over ultimately periodic $\mathcal{TEL}^{\bigcirc}$ TBoxes is **PSpace**-complete in data complexity

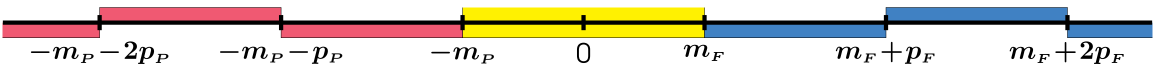
Ultimate Periodicity

a quasimodel is **ultimately p-periodic** if

for each trace π, there are positive integers $m_P, p_P, m_F, p_F \leq p$ such that

$\pi_B(n - p_P) = \pi_B(n)$, for all $n \leq -m_P$

$\pi_B(n + p_F) = \pi_B(n)$, for all $n \geq m_F$



Theorem TAQ answering over ultimately periodic $\mathcal{TEL}^{\circ}$ TBoxes is **PSpace**-complete in data complexity

rewriting into DATALOG$_{1S}$
an extension of datalog with one successor function [Chomicki & Imielinski, 1988]

terms $t + i$ and $t - i$
for a 'temporal variable' t and i is a non-negative integer constant
(encoded in unary)

DATALOG$_{1S}$ is **PSpace**-complete in data complexity
and **ExpTime**-complete in combined complexity

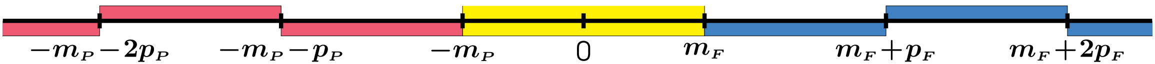
Ultimate Periodicity

a quasimodel is **ultimately p-periodic** if

for each trace π, there are positive integers $m_P, p_P, m_F, p_F \leq p$ such that

$\pi_B(n - p_P) = \pi_B(n)$, for all $n \leq -m_P$

$\pi_B(n + p_F) = \pi_B(n)$, for all $n \geq m_F$



Theorem TAQ answering over ultimately periodic $\mathcal{TEL}^{\bigcirc}$ TBoxes is **PSpace**-complete in data complexity

rewriting into DATALOG$_{1S}$

an extension of datalog with one successor function [Chomicki & Imielinski, 1988]

terms $t + i$ and $t - i$

for a ‘temporal variable’ t and i is a non-negative integer constant (encoded in unary)

DATALOG$_{1S}$ is **PSpace**-complete in data complexity and **ExpTime**-complete in combined complexity

are all $\mathcal{TEL}^{\bigcirc}$ TBoxes ultimately periodic?

Restricted Use of Rigid Roles

Theorem TAQ answering over $\mathcal{TEL}^{\bigcirc}$ **without rigid roles** is **PSpace**-complete in combined and **PTime**-complete in data complexity

observation: traces are ultimately periodic with the 'prefix' $|\mathcal{A}| + 2^{O(|\mathcal{T}|)}$ and period $2^{O(|\mathcal{T}|)}$

Restricted Use of Rigid Roles

Theorem TAQ answering over $\mathcal{TEL}^{\bigcirc}$ **without rigid roles** is **PSpace**-complete in combined and **PTime**-complete in data complexity

observation: traces are ultimately periodic with the 'prefix' $|\mathcal{A}| + 2^{O(|\mathcal{T}|)}$ and period $2^{O(|\mathcal{T}|)}$

Theorem TAQ answering over $\mathcal{TEL}^{\bigcirc}$ **without rigid roles on the right of CIs** is in **ExpTime** in combined and **PSpace**-complete in data complexity

proof: rewriting into DATALOG_{1S}

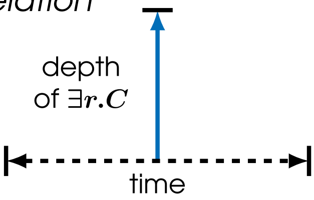
Standard Acyclicity

acyclic TBoxes: definitions $A \equiv C$ (the relation 'defined by' is acyclic)

Theorem TAQ answering over **acyclic** $\mathcal{TEL}^{\circ}$ is in LogTime-uniform **AC^0** in data and in **PTime** in combined complexity

proof: rewriting into FO with *one successor relation*

acyclicity restricts **both** dimensions

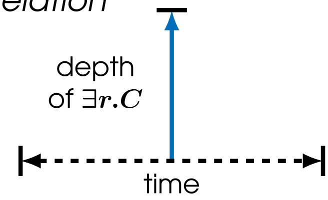


Standard Acyclicity

acyclic TBoxes: definitions $A \equiv C$ (the relation 'defined by' is acyclic)

Theorem TAQ answering over **acyclic** $\mathcal{TEL}^{\circ}$ is in LogTime-uniform **AC^0** in data and in **PTime** in combined complexity

proof: rewriting into FO with *one successor relation*



acyclicity restricts **both** dimensions

– rigid concepts are not expressible

– no recurring patterns

Standard Acyclicity

acyclic TBoxes: definitions $A \equiv C$ (the relation 'defined by' is acyclic)

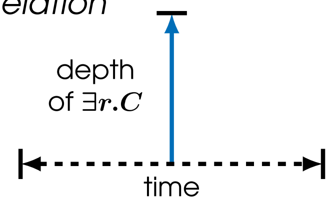
Theorem TAQ answering over **acyclic** $\mathcal{TEL}^{\bigcirc}$ is in LogTime-uniform **AC^0** in data and in **PTime** in combined complexity

proof: rewriting into FO with *one successor relation*

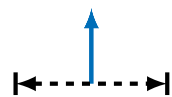
acyclicity restricts **both** dimensions

– rigid concepts are not expressible

– no recurring patterns

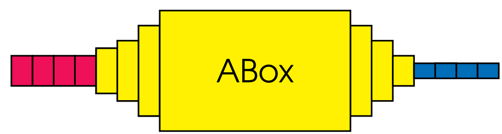


Temporal Acyclicity



Theorem TAQ answering over temporally acyclic $\mathcal{TEL}^{\bigcirc}$ (with rigid concepts) is **PTime**-complete in data and combined complexity

$\bigcirc_P A \sqsubseteq B$ or $\bigcirc_F B \sqsubseteq A$ $\Longrightarrow$
the 'rank' of B = the rank of A + 1



Standard Acyclicity

acyclic TBoxes: definitions $A \equiv C$ (the relation ‘defined by’ is acyclic)

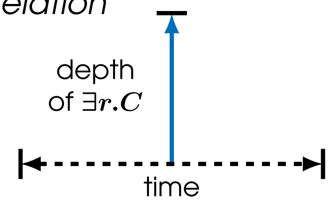
Theorem TAQ answering over **acyclic** $\mathcal{TEL}^{\bigcirc}$ is in LogTime-uniform $\mathbf{AC}^0$ in data and in **PTime** in combined complexity

proof: rewriting into FO with *one successor relation*

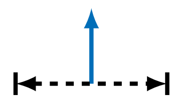
acyclicity restricts **both** dimensions

– rigid concepts are not expressible

– no recurring patterns



Temporal Acyclicity

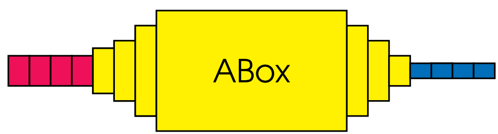


Theorem TAQ answering over temporally acyclic $\mathcal{TEL}^{\bigcirc}$ (with rigid concepts) is **PTime**-complete in data and combined complexity

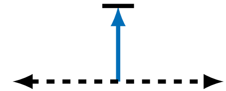
$\bigcirc_P A \sqsubseteq B$ or $\bigcirc_F B \sqsubseteq A \implies$

the ‘rank’ of B = the rank of A + 1

NB: contains full $\mathcal{EL}$



DL Acyclicity



Theorem TAQ answering over DL-acyclic $\mathcal{TEL}^{\bigcirc}$ TBoxes of depth $k \geq 1$ is k-ExpSpace-complete in combined complexity and NC^1-complete in data complexity

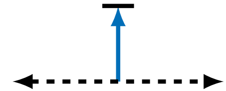
DL Acyclicity

Theorem TAQ answering over DL-acyclic $\mathcal{TEL}^{\bigcirc}$ TBoxes of depth $k \geq 1$ is k-ExpSpace-complete in combined complexity and NC^1-complete in data complexity

$A \sqsubseteq \exists r.B \quad \Longrightarrow \quad$ the rank of $B >$ the rank of A

$+$ bound on the ranks $\quad \Longrightarrow \quad$ ultimately periodic (and so, in k-ExpSpace)
the matching lower bound is by ‘yardsticks’

DL Acyclicity



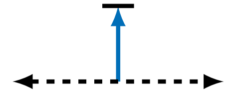
Theorem TAQ answering over DL-acyclic $\mathcal{TEL}^{\bigcirc}$ TBoxes of depth $k \geq 1$ is k-**ExpSpace**-complete in combined complexity and **NC**^{1}-complete in data complexity

$A \sqsubseteq \exists r.B \quad \Longrightarrow \quad$ the rank of $B >$ the rank of A

$+$ bound on the ranks $\quad \Longrightarrow \quad$ ultimately periodic (and so, in k-ExpSpace)
the matching lower bound is by 'yardsticks'

NC^1 comes from the temporal component alone

NB: contains the full Horn-LTL and can express rigid concepts



DL Acyclicity

Theorem TAQ answering over DL-acyclic $\mathcal{TEL}^{\bigcirc}$ TBoxes of depth $k \geq 1$ is k-ExpSpace-complete in combined complexity and NC^1-complete in data complexity

$A \sqsubseteq \exists r.B \quad \Longrightarrow \quad$ the rank of $B >$ the rank of A

$+$ bound on the ranks $\quad \Longrightarrow \quad$ ultimately periodic (and so, in k-ExpSpace)
the matching lower bound is by 'yardsticks'

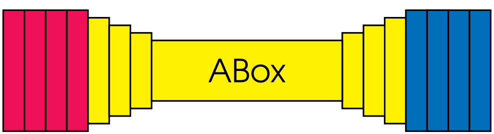
NC^1 comes from the temporal component alone

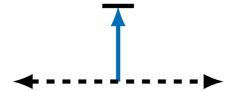
NB: contains the full Horn-LTL and can express rigid concepts

Inflationary $\mathcal{TEL}$

$\mathcal{TEL}^{\Diamond}_{\text{infl}}$ is $\mathcal{TEL}$ without $\bigcirc_*$ but with $\Diamond_*$ on the **left-hand side** of CIs only
similar to TQL for DL-Lite [Artale et al., 2013] and inflationary Datalog$_{1S}$ [Chomicki, 1990]

Theorem TAQ answering over $\mathcal{TEL}^{\Diamond}_{\text{infl}}$ is **PTime**-complete in both data and combined complexity





DL Acyclicity

Theorem TAQ answering over DL-acyclic $\mathcal{TEL}^{\bigcirc}$ TBoxes of depth $k \geq 1$ is k-ExpSpace-complete in combined complexity and NC^1-complete in data complexity

$A \sqsubseteq \exists r.B \quad \Longrightarrow \quad$ the rank of $B >$ the rank of A

$+$ bound on the ranks $\quad \Longrightarrow \quad$ ultimately periodic (and so, in k-ExpSpace)
the matching lower bound is by 'yardsticks'

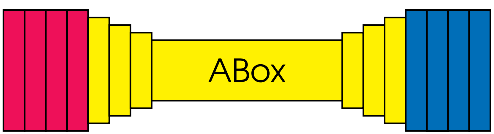
NC^1 comes from the temporal component alone

NB: contains the full Horn-LTL and can express rigid concepts

Inflationary $\mathcal{TEL}$

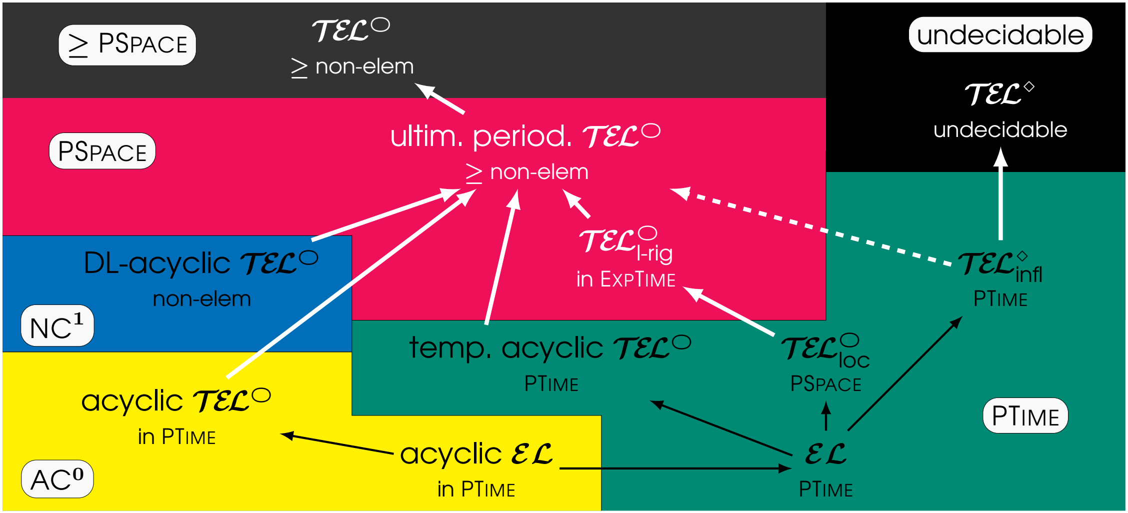
$\mathcal{TEL}^{\Diamond}_{\text{infl}}$ is $\mathcal{TEL}$ without $\bigcirc_*$ but with $\Diamond_*$ on the **left-hand side** of CIs only
similar to TQL for DL-Lite [Artale et al., 2013] and inflationary Datalog$_{1S}$ [Chomicki, 1990]

Theorem TAQ answering over $\mathcal{TEL}^{\Diamond}_{\text{infl}}$ is **PTime**-complete in both data and combined complexity



NB: can express rigid concepts $\Diamond_P \Diamond_F C \sqsubseteq C$

Summary



- full $\mathcal{TEL}^{\bigcirc}$
- conjunctive queries
- binary representation of time instants / interval encoding, e.g., $A(a, [n_1, n_2])$