Preserving Constraints with the Stable Chase

David Carral    Markus Krötzsch    Maximilian Marx    Ana Ozaki
Sebastian Rudolph

TU Dresden

2018-03-27
ICDT 2018

Full paper: https://iccl.inf.tu-dresden.de/web/Incidentals/en
A bit of background

Motivation

- ontological modelling on knowledge graphs using *tuple-generating dependencies (TGDs)*: $\varphi(x, y) \rightarrow \exists z. \psi(x, z)$
- Ontologies for Knowledge Graphs: Breaking the Rules, Krötzsch & Thost [ISWC 2016]
Motivation

- ontological modelling on knowledge graphs using *tuple-generating dependencies* (TGDs): $\varphi(x, y) \rightarrow \exists z. \psi(x, z)$
- Ontologies for Knowledge Graphs: Breaking the Rules, Krötzsch & Thost [ISWC 2016]

Observation

- Some functional dependencies may not hold, but assuming them doesn’t change any (boolean) query answers.
- $\leadsto$ “incidental functional dependencies”
- Query rewriting becomes easier if (some) incidentals are known.
A bit of background

Motivation

- ontological modelling on knowledge graphs using *tuple-generating dependencies (TGDs)*: $\varphi(x, y) \rightarrow \exists z. \psi(x, z)$
- Ontologies for Knowledge Graphs: Breaking the Rules, Krötzsch & Thost [ISWC 2016]

Observation

- Some functional dependencies may not hold, but assuming them doesn’t change any (boolean) query answers.
- $\leadsto$ “incidental functional dependencies”
- Query rewriting becomes easier if (some) incidentals are known.

Now

- What about *incidental TGDs*?
Incidental TGDs

Example

Consider $\Sigma = \{\exists x, y. R(x, y), \ R(x, y) \rightarrow \exists z. R(y, z)\}$. Then

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots$$

(1)

is the unique universal model.
Incidental TGDs

Example

Consider $\Sigma = \{ \exists x, y. R(x, y), \quad R(x, y) \rightarrow \exists z. R(y, z) \}$. Then

\[
\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]

is the unique universal model. Add $\rho = R(y, z) \rightarrow \exists x. R(x, y)$:

\[
\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]
Incidental TGDs

Example

Consider $\Sigma = \{\exists x, y. R(x, y), \quad R(x, y) \rightarrow \exists z. R(y, z)\}$. Then

\[
\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]

(1)

is the unique universal model. Add $\rho = R(y, z) \rightarrow \exists x. R(x, y)$:

\[
\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]

(2)

- Clearly, $\Sigma \not\models \rho$. 

Definition

A TGD $\rho$ is incidental for a set $\Sigma$ of TGDs if $\text{BCQ}(\Sigma) = \text{BCQ}(\Sigma \cup \{\rho\})$.

$\text{ICDT}(\Sigma)$ is the set of all TGDs incidental for $\Sigma$. 

Maximilian Marx (TU Dresden)
Incidental TGDs

Example

Consider $\Sigma = \{\exists x, y. R(x, y), \ R(x, y) \rightarrow \exists z. R(y, z)\}$. Then

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots$$

(1)

is the unique universal model. Add $\rho = R(y, z) \rightarrow \exists x. R(x, y)$:

$$\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots$$

(2)

- Clearly, $\Sigma \not\models \rho$.
- Is there a boolean conjunctive query (BCQ) separating the two models?
### Incidental TGDs

**Example**

Consider \( \Sigma = \{ \exists x, y. R(x, y), \quad R(x, y) \rightarrow \exists z. R(y, z) \} \). Then

\[
\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]

is the unique universal model. Add \( \rho = R(y, z) \rightarrow \exists x. R(x, y) \):

\[
\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots
\]

- Clearly, \( \Sigma \not\models \rho \).
- Is there a *boolean conjunctive query (BCQ)* separating the two models?
- No, \( \text{BCQ}(\Sigma) = \text{BCQ}(\Sigma \cup \{ \rho \}) \).
Incidental TGDs

Example

Consider $\Sigma = \{\exists x, y. R(x, y), R(x, y) \rightarrow \exists z. R(y, z)\}$. Then

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots$$

is the unique universal model. Add $\rho = R(y, z) \rightarrow \exists x. R(x, y)$:

$$\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots$$

Clearly, $\Sigma \not\models \rho$.

Is there a boolean conjunctive query (BCQ) separating the two models?

No, $\text{BCQ}(\Sigma) = \text{BCQ}(\Sigma \cup \{\rho\})$.

Definition

A TGD $\rho$ is *incidental* for a set $\Sigma$ of TGDs if $\text{BCQ}(\Sigma) = \text{BCQ}(\Sigma \cup \{\rho\})$. $\text{ICDT}(\Sigma)$ is the set of all TGDs incidental for $\Sigma$. 

The finite case

**Definition**

**INCIDENTAL**: Given $\Sigma$ set of TGDs and $\rho$ TGD, decide whether $\rho \in \text{ICDT}(\Sigma)$.

**Theorem**

Let $\Sigma$ be a set of TGDs with finite universal model $I$. Then $\rho \in \text{ICDT}(\Sigma)$ iff $\text{core } I| = \rho$. Use the core chase to compute $\text{core } I| = \rho \Rightarrow$ Incidental is decidable if $\Sigma$ has a finite universal model. But what can we say in general?
**The finite case**

**Definition**

**INCIDENTAL**: Given $\Sigma$ set of TGDs and $\rho$ TGD, decide whether $\rho \in \text{ICDT}(\Sigma)$.

**Theorem**

Let $\Sigma$ be a set of TGDs with finite universal model $\mathcal{I}$. Then $\rho \in \text{ICDT}(\Sigma)$ iff $\text{core} \mathcal{I} \models \rho$.

Use the core chase to compute $\text{core} \mathcal{I}$.
The finite case

**Definition**

**INCIDENTAL:** Given $\Sigma$ set of TGDs and $\rho$ TGD, decide whether $\rho \in \text{ICDT}(\Sigma)$.

**Theorem**

*Let $\Sigma$ be a set of TGDs with finite universal model $\mathcal{I}$. Then $\rho \in \text{ICDT}(\Sigma)$ iff $\text{core}\ \mathcal{I} \models \rho$.***

- Use the *core chase* to compute $\text{core}\ \mathcal{I}$
- check $\text{core}\ \mathcal{I} \models \rho$
- $\rightsquigarrow$ **INCIDENTAL** is decidable if $\Sigma$ has a finite universal model.
The finite case

**Definition**

**INCIDENTAL**: Given $\Sigma$ set of TGDs and $\rho$ TGD, decide whether $\rho \in \text{ICDT}(\Sigma)$.

**Theorem**

Let $\Sigma$ be a set of TGDs with finite universal model $\mathcal{I}$. Then $\rho \in \text{ICDT}(\Sigma)$ iff $\text{core } \mathcal{I} \models \rho$.

- Use the *core chase* to compute $\text{core } \mathcal{I}$
- check $\text{core } \mathcal{I} \models \rho$
- $\rightsquigarrow$ **INCIDENTAL** is decidable if $\Sigma$ has a finite universal model.
- But what can we say in general?
Deciding INCIDENTAL

**Theorem**

INCIDENTAL is $\Pi_2^0$-complete, and thus neither in RE nor in coRE.
Deciding INCIDENTAL?

**Theorem**

*INCIDENTAL is $\Pi^0_2$-complete, and thus neither in RE nor in coRE.*

- Can we do better if BCQ entailment is decidable?

Maximilian Marx (TU Dresden)
Deciding **INCIDENTAL**?

**Theorem**

**INCIDENTAL** is $\Pi^0_2$-complete, and thus neither in RE nor in coRE.

- Can we do better if BCQ entailment is decidable?

**Theorem**

Let $\mathcal{C}$ be a class of sets of TGDs for which BCQ entailment is decidable. Then **INCIDENTAL** is in coRE for any $\Sigma \in \mathcal{C}$. 

Deciding **INCIDENTAL**?

**Theorem**

**INCIDENTAL** is \( \Pi^0_2 \)-complete, and thus neither in \( \text{RE} \) nor in \( \text{coRE} \).

- Can we do better if BCQ entailment is decidable?

**Theorem**

Let \( \mathcal{C} \) be a class of sets of TGDs for which BCQ entailment is decidable. Then **INCIDENTAL** is in \( \text{coRE} \) for any \( \Sigma \in \mathcal{C} \).

- Idea: if \( \rho \notin \text{ICDT}(\Sigma) \), there is a BCQ \( q \) with \( \Sigma \not\models q \) and \( \Sigma \cup \{\rho\} \models q \).
- Can we do better?
Deciding **INCIDENTAL**?

**Theorem**

**INCIDENTAL** is $\Pi^0_2$-complete, and thus neither in RE nor in coRE.

- Can we do better if BCQ entailment is decidable?

**Theorem**

Let $\mathcal{C}$ be a class of sets of TGDs for which BCQ entailment is decidable. Then **INCIDENTAL** is in coRE for any $\Sigma \in \mathcal{C}$.

- Idea: if $\rho \notin \text{ICDT}(\Sigma)$, there is a BCQ $q$ with $\Sigma \not\models q$ and $\Sigma \cup \{\rho\} \models q$.
- Can we do better?
- Unfortunately, no.
Undecidability of INCIDENTAL

Theorem

There is a class \( \mathcal{C} \) of sets of TGDs and a full dependency \( \rho \) such that

- BCQ entailment for \( \Sigma \in \mathcal{C} \) is decidable,
- \( \Sigma \cup \{\rho\} \in \mathcal{C} \) for any \( \Sigma \in \mathcal{C} \), and
- checking \( \rho \in \text{ICDT}(\Sigma) \) is undecidable.
There is a class $C$ of sets of TGDs and a full dependency $\rho$ such that
- BCQ entailment for $\Sigma \in C$ is decidable,
- $\Sigma \cup \{\rho\} \in C$ for any $\Sigma \in C$, and
- checking $\rho \in \text{ICDT}(\Sigma)$ is undecidable.

What about incidentals in general?
- Recall: in the finite case, $\rho \in \text{ICDT}(\Sigma)$ iff core $\mathcal{I} \models \rho$
- Is there some universal model that entails all incidental TGDs?
- The core looks like a promising candidate.
Cores of infinite instances

Different definitions of core agree on finite instances, but differ in general.

**Definition**

An instance \( \mathcal{I} \) is a core if every endomorphism \( h : \mathcal{I} \to \mathcal{I} \) is an embedding.

A core \( \mathcal{I} \) is a core of \( \mathcal{J} \) if \( \mathcal{I} = \mathcal{J}|_{h(\mathcal{J})} \) for an endomorphism \( h : \mathcal{J} \to \mathcal{J} \).
Cores of infinite instances

Different definitions of core agree on finite instances, but differ in general.

**Definition**

An instance $\mathcal{I}$ is a core if every endomorphism $h : \mathcal{I} \rightarrow \mathcal{I}$ is an embedding. A core $\mathcal{I}$ is a core of $\mathcal{J}$ if $\mathcal{I} = \mathcal{J} \mid_{h(\mathcal{J})}$ for an endomorphism $h : \mathcal{J} \rightarrow \mathcal{J}$.

Unfortunately, there are sets $\Sigma$ of TGDs with universal models $\mathcal{I}$ such that

- $\mathcal{I}$ doesn’t have a core,
- $\mathcal{I}$ has two non-isomorphic cores, or
- $\mathcal{I}$ has a core that is not a model of $\Sigma$. 

Theorem

Let $\Sigma$ be a set of TGDs. There is a core $\mathcal{I}$ with $\mathcal{I} \mid_{BCQ(\mathcal{I})} = \Sigma$, $BCQ(\mathcal{I}) = BCQ(\Sigma)$, and $\rho \in IC\mathcal{D}\mathcal{T}(\Sigma)$ iff $\mathcal{I} \mid_{h(\mathcal{J})}$. 

Can we generalise the core chase to this setting?
Cores of infinite instances

Different definitions of core agree on finite instances, but differ in general.

**Definition**
An instance $\mathcal{I}$ is a core if every endomorphism $h : \mathcal{I} \to \mathcal{I}$ is an embedding. A core $\mathcal{I}$ is a core of $\mathcal{J}$ if $\mathcal{I} = \mathcal{J}|_{h(\mathcal{J})}$ for an endomorphism $h : \mathcal{J} \to \mathcal{J}$.

Unfortunately, there are sets $\Sigma$ of TGDs with universal models $\mathcal{I}$ such that
- $\mathcal{I}$ doesn’t have a core,
- $\mathcal{I}$ has two non-isomorphic cores, or
- $\mathcal{I}$ has a core that is not a model of $\Sigma$.

**Theorem**
Let $\Sigma$ be a set of TGDs. There is a core $\mathcal{I}$ with $\mathcal{I} \models \Sigma$, $BCQ(\mathcal{I}) = BCQ(\Sigma)$, and $\rho \in ICDT(\Sigma)$ iff $\mathcal{I} \models \rho$. 
Cores of infinite instances

Different definitions of core agree on finite instances, but differ in general.

**Definition**

An instance \( \mathcal{I} \) is a core if every endomorphism \( h : \mathcal{I} \to \mathcal{I} \) is an embedding. A core \( \mathcal{I} \) is a core of \( \mathcal{J} \) if \( \mathcal{I} = \mathcal{J}|_{h(\mathcal{J})} \) for an endomorphism \( h : \mathcal{J} \to \mathcal{J} \).

Unfortunately, there are sets \( \Sigma \) of TGDs with universal models \( \mathcal{I} \) such that
- \( \mathcal{I} \) doesn’t have a core,
- \( \mathcal{I} \) has two non-isomorphic cores, or
- \( \mathcal{I} \) has a core that is not a model of \( \Sigma \).

**Theorem**

Let \( \Sigma \) be a set of TGDs. There is a core \( \mathcal{I} \) with \( \mathcal{I} \models \Sigma \), \( \text{BCQ}(\mathcal{I}) = \text{BCQ}(\Sigma) \), and \( \rho \in \text{ICDT}(\Sigma) \) iff \( \mathcal{I} \models \rho \).

- Can we generalise the core chase to this setting?
The stable chase by example

\[ \Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \to \exists x. R(x, y), \]
\[ R(x, y) \land S(x, y) \to \exists z. R(y, z) \land S(y, z) \} \]
The stable chase by example

\[\Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \]
\[R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \}\]

Q_1

n_0  n_1
The stable chase by example

\[ \Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \]
\[ R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \} \]
The stable chase by example

$$\Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \\
R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \}$$

$Q_1$

$Q_2$

$Q_3$
The stable chase by example

\[ \Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \]
\[ R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \} \]
The stable chase by example

\[ \Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \]
\[ \quad R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \} \]
The stable chase by example

\[ \Sigma = \{ \exists x, y. R(x, y) \land S(x, y), \quad R(y, z) \rightarrow \exists x. R(x, y), \]
\[ \quad R(x, y) \land S(x, y) \rightarrow \exists z. R(y, z) \land S(y, z) \} \]
The stable chase

The idea

- Consider a sequence of initial segments of chase sequences (prefixes):
  - apply TGDs to last instance and obtain a longer prefix, or
  - rewrite prefix according to some non-embedding endomorphism between instances \( \leadsto \) enforce an embedding
  - instances can only be rewritten finitely often and stabilise at some point
  - The *stable chase* is the union of all stable instances.
The stable chase

The idea

- Consider a sequence of initial segments of chase sequences (prefixes):
  - apply TGDs to last instance and obtain a longer prefix, or
  - *rewrite* prefix according to some non-embedding endomorphism between instances $\leadsto$ enforce an embedding
  - instances can only be rewritten finitely often and stabilise at some point
  - The *stable chase* is the union of all stable instances.

Theorem

For any set $\Sigma$ of TGDs, the stable chase $\mathcal{I}$ of $\Sigma$ is a core with $\mathcal{I} \models \Sigma$, $\text{BCQ}(\mathcal{I}) = \text{BCQ}(\Sigma)$, and $\rho \in \text{ICDT}(\Sigma)$ iff $\mathcal{I} \models \rho$ for any full dependency $\rho$. 

Beware: Stability of an instance is undecidable.
The stable chase

The idea

- Consider a sequence of initial segments of chase sequences (prefixes):
  - apply TGDs to last instance and obtain a longer prefix, or
  - rewrite prefix according to some non-embedding endomorphism between instances \(\sim\) enforce an embedding
  - instances can only be rewritten finitely often and stabilise at some point
  - The **stable chase** is the union of all stable instances.

Theorem

*For any set \(\Sigma\) of TGDs, the stable chase \(\mathcal{I}\) of \(\Sigma\) is a core with \(\mathcal{I} \models \Sigma\), \(BCQ(\mathcal{I}) = BCQ(\Sigma)\), and \(\rho \in ICDT(\Sigma)\) iff \(\mathcal{I} \models \rho\) for any full dependency \(\rho\).*

- If \(\Sigma\) has a finite universal model \(\mathcal{J}\), then core \(\mathcal{J} = \mathcal{I}\).
- Beware: Stability of an instance is undecidable.
Conclusion & Outlook

Results

- Incidence is $\Pi_2^0$-complete, and still not in RE even when BCQ entailment is decidable.
- Each set $\Sigma$ of TGDs has a BCQ-equivalent core $I$ with $I \models \text{ICDT}(\Sigma)$.
- The stable chase generalises the core chase to classes that don’t admit finite universal models.
- The stable chase yields a core that characterises the full incidental dependencies.

Future work

- Further generalise the stable chase to characterise all incidentals.
- Investigate complexity of Incidental for decidable classes.
- Design (incomplete) algorithms that compute incidentals.