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*): $\varphi(\boldsymbol{x}, \boldsymbol{y}) \to \exists \boldsymbol{z}.\ \psi(\boldsymbol{x}, \boldsymbol{z})$
- Ontologies for Knowledge Graphs: Breaking the Rules, Krötzsch & Thost [ISWC 2016]

A bit of background

Motivation

- ontological modelling on knowledge graphs using *tuple-generating dependencies* (*TGDs*): $\varphi(\boldsymbol{x}, \boldsymbol{y}) \to \exists \boldsymbol{z}.\ \psi(\boldsymbol{x}, \boldsymbol{z})$
- 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.
- $\rightsquigarrow$ "*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*): $\varphi(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}.\ \psi(\boldsymbol{x}, \boldsymbol{z})$
- 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.
- $\rightsquigarrow$ "*incidental functional dependencies*"
- Query rewriting becomes easier if (some) incidentals are known.

Now

- What about *incidental TGDs*?

Incidental TGDs

Example

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{1}$$

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{1}$$

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

$$\dots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{2}$$

Incidental TGDs

Example

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{1}$$

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

$$\dots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{2}$$

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

Incidental TGDs

Example

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{1}$$

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

$$\dots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{2}$$

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

Incidental TGDs

Example

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots \tag{1}$$

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

$$\ldots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \ldots \tag{2}$$

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

Incidental TGDs

Example

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

$$\bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{1}$$

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

$$\dots \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \bullet \xrightarrow{R} \dots \tag{2}$$

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

Definition

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

The finite case

Definition

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

The finite case

Definition

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

Theorem

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

The finite case

Definition

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

Theorem

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

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

The finite case

Definition

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

Theorem

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

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

Deciding INCIDENTAL

Theorem

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

Deciding INCIDENTAL?

Theorem

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

- Can we do better if BCQ entailment is decidable?

Deciding INCIDENTAL?

Theorem

INCIDENTAL *is* Π^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* Π^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 $\rho \notin \mathsf{ICDT}(\Sigma)$, 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* Π^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 ρ 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 $\rho \in \mathsf{ICDT}(\Sigma)$ is undecidable.*

Undecidability of $\textsc{Incidental}$

Theorem

There is a class $\mathcal{C}$ of sets of TGDs and a full dependency ρ 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 $\rho \in \mathsf{ICDT}(\Sigma)$ is undecidable.*

What about incidentals in general?

- Recall: in the finite case, $\rho \in \mathsf{ICDT}(\Sigma)$ 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 $h : \mathcal{I} \to \mathcal{I}$ 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 Σ 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 Σ.

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 Σ 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 Σ.

Theorem

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

Theorem

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

- Can we generalise the core chase to this setting?

The stable chase by example

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

The stable chase by example

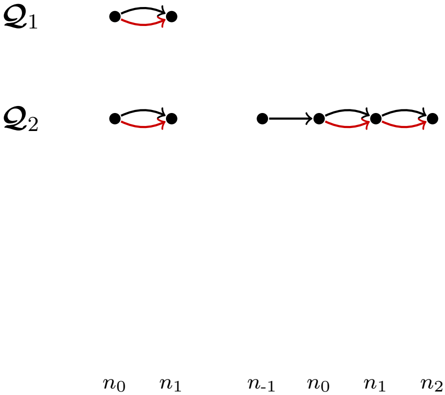
$$\Sigma = \{\exists x, y.\ R(x,y) \wedge S(x,y), \qquad R(y,z) \to \exists x.\ R(x,y),$$
$$R(x,y) \wedge S(x,y) \to \exists z.\ R(y,z) \wedge S(y,z)\}$$

$\mathcal{Q}_1$

n_0 n_1

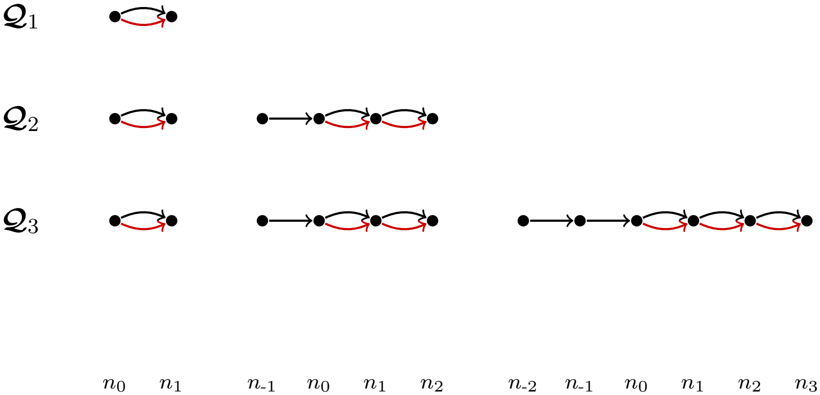
The stable chase by example

$$\Sigma = \{\exists x, y.\ R(x,y) \wedge S(x,y), \qquad R(y,z) \to \exists x.\ R(x,y),$$
$$R(x,y) \wedge S(x,y) \to \exists z.\ R(y,z) \wedge S(y,z)\}$$



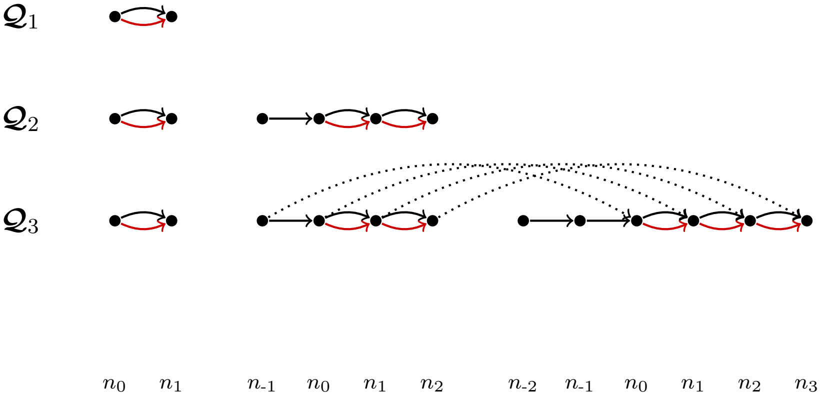
The stable chase by example

$$\Sigma = \{\exists x, y.\ R(x,y) \wedge S(x,y), \qquad R(y,z) \to \exists x.\ R(x,y),$$
$$R(x,y) \wedge S(x,y) \to \exists z.\ R(y,z) \wedge S(y,z)\}$$



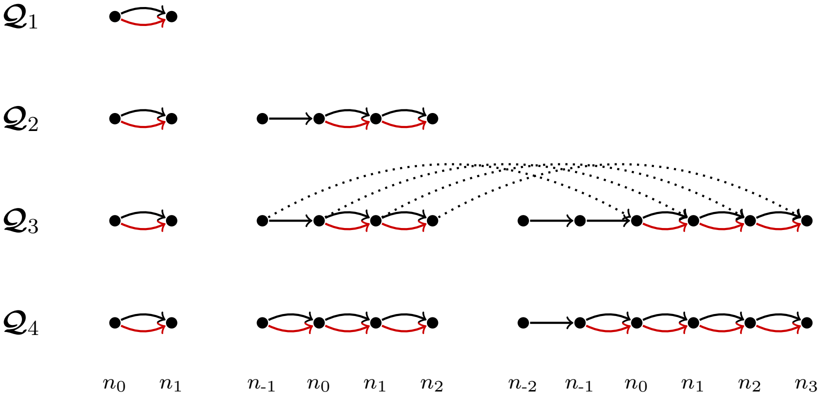
The stable chase by example

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



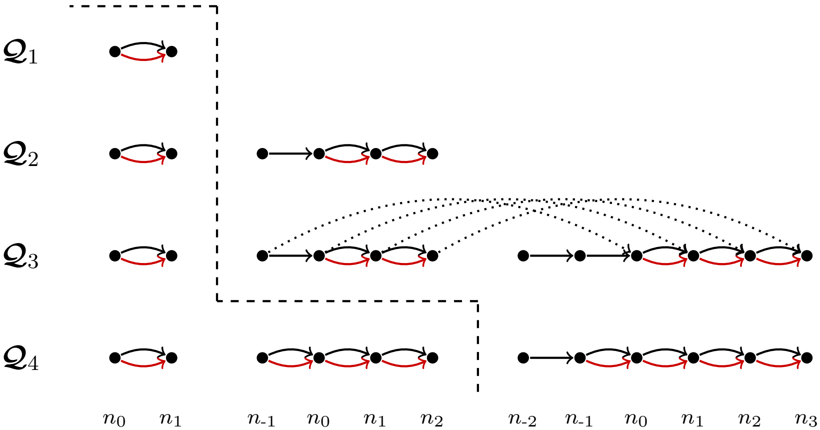
The stable chase by example

$$\Sigma = \{\exists x, y.\ R(x,y) \wedge S(x,y), \qquad R(y,z) \to \exists x.\ R(x,y),$$
$$R(x,y) \wedge S(x,y) \to \exists z.\ R(y,z) \wedge S(y,z)\}$$



The stable chase by example

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



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

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

- If Σ has a finite universal model $\mathcal{J}$, then core $\mathcal{J} = \mathcal{I}$.
- Beware: Stability of an instance is undecidable.

Conclusion & Outlook

Results

- Incidentality is Π_2^0-complete, and still not in RE even when BCQ entailment is decidable.
- Each set Σ of TGDs has a BCQ-equivalent core $\mathcal{I}$ with $\mathcal{I} \models \mathsf{ICDT}(\Sigma)$.
- 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.