CS 595 - Hot topics in database systems:
Data Provenance

I. Database Provenance

I.1 Provenance Models and Systems

Boris Glavic

October 10, 2012

Outline

Set Difference

Recall

- $[[R - S]] = \{t \mid t \in R \wedge t \notin S\}$
- $[[R - S]] = \{t^{n-m} \mid t^n \in R \wedge t^m \in S\}$

Properties

ILLINOIS INSTITUTE OF TECHNOLOGY

Set Difference

Recall

- $[[R - S]] = \{t \mid t \in R \wedge t \notin S\}$
- $[[R - S]] = \{t^{n-m} \mid t^n \in R \wedge t^m \in S\}$

Properties

- Negation

ILLINOIS INSTITUTE OF TECHNOLOGY

Set Difference

Recall

- $[[R - S]] = \{t \mid t \in R \wedge t \notin S\}$
- $[[R - S]] = \{t^{n-m} \mid t^n \in R \wedge t^m \in S\}$

Properties

- Negation
- $\Rightarrow$Non-monotone
  - Add tuples to the input may remove tuples from the output

ILLINOIS INSTITUTE OF TECHNOLOGY

Set Difference

Recall

- $[[R - S]] = \{t \mid t \in R \wedge t \notin S\}$
- $[[R - S]] = \{t^{n-m} \mid t^n \in R \wedge t^m \in S\}$

Properties

- Negation
- $\Rightarrow$Non-monotone
  - Add tuples to the input may remove tuples from the output
- $\Rightarrow$Non-transitive provenance

ILLINOIS INSTITUTE OF TECHNOLOGY

Non-Transitive Provenance

Implies

1. Provenance no longer local

Non-Transitive Provenance

Implies

1. Provenance no longer local

Example

$q_1 = S - T$ $\qquad$ $q_2 = R - (S - T)$

- Take set-semantics
- For q_1, tuples from T never belong to provenance
- For q_2, tuple from T can belong to provenance

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

Non-Transitive Provenance

Implies

1. Provenance no longer local
   - ⇒Problematic for approaches that rely on transitivity
   - ⇒Compositional rules only work if negation in provenance representation

Example

$$q_1 = S - T \qquad\qquad q_2 = R - (S - T)$$

- t_1 does not belong to provenance of q_1
- ⇒t_1 not in provenance of q_2 (transitivity)!

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

Non-Transitive Provenance

Implies

1. Provenance no longer local
2. Provenance of non-exiting tuple belongs to result provenance

Example

$$q_1 = S - T \qquad q_2 = R - (S - T)$$

- q_1 returns empty result
- Still inputs of q_1 belong to provenance

R

	a
r_1	1

S

	b
s_1	1

T

	b
t_1	1
t_2	3

Q_1

a

Q_2

	a
t	1

Why-Provenance

Why-Provenance

- Why-provenance
  - Syntactic definition ⇒undefined
- Insensitive Why-provenance
  - Minimality ⇒no-tuples from right-hand side tree belong to provenance

Why-Provenance

Example

$$IWhy(q_1, t) = \{\{s_2\}\} \qquad IWhy(q_2, t) = \{\{r_1\}\}$$

$$q_1 = S - T \qquad q_2 = R - (S - T)$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

ILLINOIS INSTITUTE OF TECHNOLOGY

Lineage

Lineage

- Apply declarative definition + transitivity
- Recall compositional rule
  - $Lin((R - S), t) =< Lin(R, t), \sigma_{\mathbf{S} \neq t}(S) >$

Lineage

Example

$Lin(q_1, t) =< \{s_2\}, \{t_1, t_2\} >$ $Lin(q_2, t) =< \{r_1\}, \{s_2\}, \{t_1, t_2\} >$

$q_1 = S - T$ $q_2 = R - (S - T)$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

ILLINOIS INSTITUTE OF TECHNOLOGY

Lineage

Example

$$Lin(q_1, t) = < \{r_1\}, \{\} > \qquad Lin(q_2, t) = < \{r_1\}, \{\}, \{\} >$$

$$q_1 = S - T \qquad q_2 = R - (S - T)$$

R

	a
r_1	1

S

	b
s_1	1

T

	b
t_1	1
t_2	3

Q_1

a

Q_2

	a
t	1

ILLINOIS INSTITUTE OF TECHNOLOGY

Provenance Polynomials

Provenance Polynomials

- Semiring algebra does not include set-difference!

ILLINOIS INSTITUTE OF TECHNOLOGY

Perm-Influence Provenance Model

Perm Influence

- Apply declarative definition
- Recall compositional rules
  - $\mathcal{PI}((R - S), t) = \{< t^n, \perp >| \ t^n \in R\}$

ILLINOIS INSTITUTE OF TECHNOLOGY

Perm-Influence Provenance Model

Example

$$\mathcal{PI}(q_1, t) = \{< s_2, \perp >\} \qquad \mathcal{PI}(q_2, t) = \{< r_1, \perp, \perp >\}$$

$$q_1 = S - T \qquad q_2 = R - (S - T)$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

Causality

Perm Influence

- Apply declarative definition
  - t' is actual cause for t iff
  - $\exists I' \subseteq (I - \{t'\}) : t \in Q(I - I') \wedge t \notin Q(I - I' - \{t'\})$
- $\Rightarrow$for single $R - S$, only tuples from R
- $\Rightarrow$for $R - (S - T)$, both R and T tuples can be counterfactual causes

Causality

Example

$$Cau(q_1, t) = \{s_2\} \qquad Cau(q_2, t) = \{r_1, t_1\}$$

$$q_1 = S - T \qquad q_2 = R - (S - T)$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
t	2

Q_2

	a
t	1

ILLINOIS INSTITUTE OF TECHNOLOGY

Summary - Dealing with Set-Difference

- Insensitive Why: NO - misses positive influence
- Lineage: NO - misses cases and irrelevante tuples in provenance
- Perm Influence: NO - misses positive influence
- Provenance Polynomials: NO - not defined
- Causality: YES (complexity may be different!)

Outline

Extending the Semiring Algebra for Set-Difference

Approach

- $\mathbb{Z}$-relations, relations annotated with integer numbers
- Consider relations annotated with more complex structures

$\mathbb{Z}$-relations

Rationale

- Annotate tuples with values from $\mathbb{Z}$
  - $\Rightarrow$Query equivalence is the same for positive algebra
- Define $(R - S)(t) = R(t) \ominus S(t)$
  - **Bag-semantics**: $\ominus_{\mathbb{N}}$ is truncating minus: $a \ominus_{\mathbb{N}} b = a - b$ if $a > b$ or 0 otherwise
  - **Set-semantics**: $\ominus_{\mathbb{B}}$ is $a \ominus_{\mathbb{B}} b = a \wedge \neg b$
  - **Z-relations**: $\ominus_{\mathbb{Z}}$ is $a \ominus_{\mathbb{Z}} b = a - b$

Example $\mathbb{Z}$-relation

Example

$\mathbb{Z}(q_1, t) = -1$ $\quad$ $\mathbb{Z}(q_2, t) = 4$

$\mathbb{N}(q_1, t) = 0$ $\quad$ $\mathbb{N}(q_2, t) = 3$

$\mathbb{B}(q_1, t) = false$ $\quad$ $\mathbb{B}(q_2, t) = true$

$r_1 = 3$ $\quad$ $s_1 = 4$ $\quad$ $t_1 = 5$

$q_1 = R - S$ $\quad$ $q_2 = R - (S - T)$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
$r_1 \ominus s_1$	1

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

GY

Discussion

Difference of $\mathbb{Z}$ and bag semantics

- $Q^{\mathbb{Z}}(t) = Q^{\mathbb{N}}(t)$ only if for every $q_1 - q_2$ in q:
  - $Q_1 \subseteq Q_2$
  - $\Rightarrow$the right-hand side of every set difference is contained in the left-hand side
- We have modelled set- and bag-semantics using annotations
  - How to represent the provenance of such queries?

ILLINOIS INSTITUTE OF TECHNOLOGY

Add Monus Operation

- Instead of semirings $(K, \oplus, \otimes, 0, 1)$ consider structures $(K, \oplus, \otimes, \ominus, 0, 1)$
- $\Rightarrow$Define set difference as $(R - S)(t) = R(t) \ominus S(t)$

GY

Add Monus Operation

- Instead of semirings $(K, \oplus, \otimes, 0, 1)$ consider structures $(K, \oplus, \otimes, \ominus, 0, 1)$
- $\Rightarrow$Define set difference as $(R - S)(t) = R(t) \ominus S(t)$
- Semiring equivalences = positive relational algebra equivalences
  - $\Rightarrow$New structure equivalences = relational algebra with set-difference equivalences

GY

Add Monus Operation

- Instead of semirings $(K, \oplus, \otimes, 0, 1)$ consider structures $(K, \oplus, \otimes, \ominus, 0, 1)$
- $\Rightarrow$Define set difference as $(R - S)(t) = R(t) \ominus S(t)$
- Semiring equivalences = positive relational algebra equivalences
  - $\Rightarrow$New structure equivalences = relational algebra with set-difference equivalences
- Provenance Polynomials = free semiring ($\mathbb{N}$ = bag, $\mathbb{B}$ = set)
  - $\Rightarrow$Find free structure for $(K, \oplus, \otimes, \ominus, 0, 1)$ and these equivalences
  - $\Rightarrow\mathbb{N}$-structure should be bag semantics
  - $\Rightarrow\mathbb{B}$-structure should be set semantics

Algebraic Equivalence for Positive Relational Algebra

$R \cup (S \cup T) \equiv (R \cup S) \cup T$	$a \oplus (b \oplus c) = (a \oplus b) \oplus c$
$R \cup \emptyset \equiv R$	$a \oplus 0 = a$
$R \cup S \equiv S \cup R$	$a \oplus b = b \oplus a$
$R \bowtie (S \bowtie T) \equiv (R \bowtie S) \bowtie T$	$a \otimes (b \otimes c) = (a \otimes b) \otimes c$
$R \bowtie \not1 \equiv R$	$a \otimes 1 = a$
$R \bowtie S = S \bowtie R a \otimes b$	$= b \otimes a$
$R \bowtie (S \cup T) \equiv (R \bowtie S) \cup (R \bowtie T)$	$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$
$R \bowtie \emptyset \equiv \emptyset$	$a \otimes 0 = 0$

ILLINOIS INSTITUTE OF TECHNOLOGY

Important Equivalences for Set-Difference

Relational Algebra

$$R - R \equiv \emptyset \qquad \text{(E1)}$$

New Structure

$$a \ominus a = 0 \qquad \text{(A1)}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Important Equivalences for Set-Difference

Relational Algebra

$$R - R \equiv \emptyset \quad \text{(E1)}$$

$$\emptyset - R \equiv \emptyset \quad \text{(E2)}$$

New Structure

$$a \ominus a = 0 \quad \text{(A1)}$$

$$0 \ominus a = 0 \quad \text{(A2)}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Important Equivalences for Set-Difference

Relational Algebra

$$R - R \equiv \emptyset \tag{E1}$$

$$\emptyset - R \equiv \emptyset \tag{E2}$$

$$R \cup (S - R) \equiv S \cup (R - S) \tag{E3}$$

New Structure

$$a \ominus a = 0 \tag{A1}$$

$$0 \ominus a = 0 \tag{A2}$$

$$a \oplus (b \ominus a) = b \oplus (a \ominus b) \tag{A3}$$

GY

Important Equivalences for Set-Difference

Relational Algebra

$$R - R \equiv \emptyset \tag{E1}$$

$$\emptyset - R \equiv \emptyset \tag{E2}$$

$$R \cup (S - R) \equiv S \cup (R - S) \tag{E3}$$

$$R - (S \cup T) \equiv (R - S) - T \tag{E4}$$

New Structure

$$a \ominus a = 0 \tag{A1}$$

$$0 \ominus a = 0 \tag{A2}$$

$$a \oplus (b \ominus a) = b \oplus (a \ominus b) \tag{A3}$$

$$a \ominus (b \oplus c) = (a \ominus b) \ominus c \tag{A4}$$

Important Equivalences for Set-Difference

Relational Algebra

$$R - R \equiv \emptyset \tag{E1}$$

$$\emptyset - R \equiv \emptyset \tag{E2}$$

$$R \cup (S - R) \equiv S \cup (R - S) \tag{E3}$$

$$R - (S \cup T) \equiv (R - S) - T \tag{E4}$$

$$R \bowtie (S - T) \equiv R \bowtie S - R \bowtie T \tag{E5}$$

New Structure

$$a \ominus a = 0 \tag{A1}$$

$$0 \ominus a = 0 \tag{A2}$$

$$a \oplus (b \ominus a) = b \oplus (a \ominus b) \tag{A3}$$

$$a \ominus (b \oplus c) = (a \ominus b) \ominus c \tag{A4}$$

$$a \otimes (b \ominus c) = a \otimes b \ominus a \otimes c \tag{A5}$$

How to Extend Semirings?

- Is there a general approach to
  - Extend semiring with $\ominus$ so that
  - . . . equivalences hold

ILLINOIS INSTITUTE OF TECHNOLOGY

How to Extend Semirings?

- Is there a general approach to
  - Extend semiring with $\ominus$ so that
  - ... equivalences hold
- YES: define $\ominus$ using $\oplus$ and an order relation

ILLINOIS INSTITUTE OF TECHNOLOGY

Natural Order

Definition (Natural Ordered Monoids)

- Monoid $(K, \oplus, 0)$
- Define $a \leq b \Leftrightarrow \exists c : a \oplus c = b$
- If $\leq$ is a partial order then K is naturally ordered

Definition (Partial Order)

$\leq$ is an order on set K iff

$$a \leq a \qquad \text{(Reflexive)}$$

$$a \leq b \wedge b \leq c \Rightarrow a \leq c \qquad \text{(Transitive)}$$

$$a \leq b \Rightarrow b \not\leq a \qquad \text{(Antisymmetric)}$$

Examples of Naturally Order Semirings

- $\mathbb{N}$: $a \leq b \Leftrightarrow \exists c : a + c = b$
  - $\Rightarrow$Unique c exists
  - $\Rightarrow$Standard order on integers
- $\mathbb{B}$: $a \leq b \Leftrightarrow \exists c : a \vee c = b$
  - $\Rightarrow$ *false* $\leq$ *true*
- $\mathbb{N}[I]$: $a \leq b \Leftrightarrow \exists c : a + c = b$
  - $x_1 \leq x_1 + {x_2}^3$: $c = {x_2}^3$
  - $\Rightarrow$Subtraction of polynomials

ILLINOIS INSTITUTE OF TECHNOLOGY

Monus Operation

Preconditions

1. K is naturally ordered semiring
2. For each $a \leq b$ there exists a smallest c with $a \oplus c = b$
3. $(K, \oplus, \otimes, \ominus, 0, 1)$ is called m-semiring

Definition (Monus Operation)

$$a \ominus b = min(\{c \mid a \leq b \oplus c\})$$

Examples Monus Operations

Set Semantics

- $\mathbb{B}$: $a \ominus b = a \wedge \neg b$

Bag Semantics

- $\mathbb{N}$: $a \ominus b = a -_t b$ (truncated minus)

Example

$$q = R - (S - T)$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Examples Monus Operations

Set Semantics

- $\mathbb{B}$: $a \ominus b = a \wedge \neg b$

Bag Semantics

- $\mathbb{N}$: $a \ominus b = a -_t b$ (truncated minus)

Example

$$q = R - (S - T)$$

$$\mathbb{N}(q, t) = 3 - (2 - 1) = 2$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Examples Monus Operations

Set Semantics

- $\mathbb{B}$: $a \ominus b = a \wedge \neg b$

Bag Semantics

- $\mathbb{N}$: $a \ominus b = a -_t b$ (truncated minus)

Example

$$q = R - (S - T)$$

$$\mathbb{B}(q, t) = true \wedge \neg(true \wedge \neg true) = true$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Provenance Polynomials with Monus

- $\mathbb{N}[I]$ with monus
- Write polynomials as sums of products (monomials)
- $x_1 \times (2 \times x_2 + x_3) \Rightarrow 2 \times x_1 \times x_2 + 1 \times x_1 \times x_3$
- Monus is truncated minus on factors

Example

$(2{x_1}^2 + 4x_2) \ominus ({x_1}^2 + x_3) = {x_1}^2 + 4x_2$

ILLINOIS INSTITUTE OF TECHNOLOGY

Example Equivalences

Example

$$q_1 = R - (S \cup T) \qquad \equiv \qquad (R - S) - T = q_2$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
$r_1 \ominus (s_1 \oplus t_1)$	1

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

OF TECHNOLOGY

Example Equivalences

Example

Evaluation in $\mathbb{N}$:

$$r_1 = 5 \qquad s_1 = 1 \qquad t_1 = 2$$

$$r_1 \ominus (s_1 \oplus t_1) = 5 - (1 + 2) \quad = 2 = \quad (5 - 1) - 2 = r_1 \ominus (s_1 \ominus t_1)$$

$$q_1 = R - (S \cup T) \qquad \equiv \qquad (R - S) - T = q_2$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
$r_1 \ominus (s_1 \oplus t_1)$	1

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Example Equivalences

Example

Evaluation in $\mathbb{B}$:

$$r_1 = t \qquad s_1 = t \qquad t_1 = t$$

$$r_1 \ominus (s_1 \oplus t_1) = t \wedge \neg(t \vee t) = f = (t \wedge \neg t) \wedge \neg t = r_1 \ominus (s_1 \ominus t_1)$$

$$q_1 = R - (S \cup T) \qquad \equiv \qquad (R - S) - T = q_2$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
$r_1 \ominus (s_1 \oplus t_1)$	1

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Example Equivalences

Example

Evaluation in $\mathbb{N}[I]$:

$$r_1 \ominus (s_1 \oplus t_1) \qquad = r_1 = \qquad r_1 \ominus (s_1 \ominus t_1)$$

$$q_1 = R - (S \cup T) \qquad \equiv \qquad (R - S) - T = q_2$$

R

	a
r_1	1

S

	b
s_1	1
s_2	2

T

	b
t_1	1
t_2	3

Q_1

	a
$r_1 \ominus (s_1 \oplus t_1)$	1

Q_2

	a
$r_1 \ominus (s_1 \ominus t_1)$	1

Is M-semiring $\mathbb{N}[I]$ Free?

- Can we evaluate other m-semiring results
  - From $\mathbb{N}[I]$ expressions
  - Like we did for semirings
- $\Rightarrow$ Is $\mathbb{N}[I]$ the free m-semiring?

Example

Counter-Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$$\mathbb{N}[I](q,t) = {r_1}^2 \ominus r_1 = {r_1}^2$$

$$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$$

$$\mathbb{B}(q,t) = (true \wedge true) \wedge \neg true = false$$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Is M-semiring $\mathbb{N}[I]$ Free?

- Can we evaluate other m-semiring results
  - From $\mathbb{N}[I]$ expressions
  - Like we did for semirings
- $\Rightarrow$Is $\mathbb{N}[I]$ the free m-semiring?

Example

Counter-Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$$\mathbb{N}[I](q,t) = {r_1}^2 \ominus r_1 = {r_1}^2$$

$$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$$

$$\mathbb{B}(q,t) = (true \wedge true) \wedge \neg true = false$$

$$eval_{\mathbb{N}}(\mathbb{N}[I](q,t)) = 2^2 = 4$$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Is M-semiring $\mathbb{N}[I]$ Free?

- Can we evaluate other m-semiring results
  - From $\mathbb{N}[I]$ expressions
  - Like we did for semirings
- $\Rightarrow$Is $\mathbb{N}[I]$ the free m-semiring?

Example

Counter-Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$\mathbb{N}[I](q,t) = {r_1}^2 \ominus r_1 = {r_1}^2$

$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$

$\mathbb{B}(q,t) = (true \wedge true) \wedge \neg true = false$

$eval_{\mathbb{N}}(\mathbb{N}[I](q,t)) = 2^2 = 4$

$eval_{\mathbb{B}}(\mathbb{N}[I](q,t)) = true \wedge true = true$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Is M-semiring $\mathbb{N}[I]$ Free?

- Can we evaluate other m-semiring results
  - From $\mathbb{N}[I]$ expressions
  - Like we did for semirings
- $\Rightarrow$Is $\mathbb{N}[I]$ the free m-semiring? NO

Example

Counter-Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$\mathbb{N}[I](q,t) = {r_1}^2 \ominus r_1 = {r_1}^2$

$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$

$\mathbb{B}(q,t) = (true \land true) \land \neg true = false$

$eval_{\mathbb{N}}(\mathbb{N}[I](q,t)) = 2^2 = 4$

$eval_{\mathbb{B}}(\mathbb{N}[I](q,t)) = true \land true = true$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

GY

Free structure for m-semirings?

Idea

- $(\mathbb{N}^{-}[I], \oplus, \otimes, \ominus, 0, 1)$
- Do not interpret operations beyond equivalences
- $\Rightarrow$Add new uninterpreted operator $\ominus$ to expressions
- $\Rightarrow$operators still "concat expressions"
- $\Rightarrow$Equivalent expressions are the same m-semiring element

ILLINOIS INSTITUTE OF TECHNOLOGY

Free structure for m-semirings?

Idea

- $(\mathbb{N}^-[I], \oplus, \otimes, \ominus, 0, 1)$
- Do not interpret operations beyond equivalences
- $\Rightarrow$Add new uninterpreted operator $\ominus$ to expressions
- $\Rightarrow$operators still "concat expressions"
- $\Rightarrow$Equivalent expressions are the same m-semiring element

Caveat

- The resulting expressions look like polynomials over $\mathbb{Z}$
- Can we interpreting them this way?

GY

Free structure for m-semirings?

Idea

- $(\mathbb{N}^{-}[I], \oplus, \otimes, \ominus, 0, 1)$
- Do not interpret operations beyond equivalences
- ⇒Add new uninterpreted operator $\ominus$ to expressions
- ⇒operators still "concat expressions"
- ⇒Equivalent expressions are the same m-semiring element

Caveat

- The resulting expressions look like polynomials over $\mathbb{Z}$
- Can we interpreting them this way?
- NO: this would be incorrect
  - $2x^2 \ominus 3x^2 = -1x^2$? NO: no negative numbers exist
  - $2x^2 \ominus x^2 = x^2$? NO: does not work in $\mathbb{B}$

Free Structure Example

Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$$\mathbb{N}^{-}[I](q,t) = {r_1}^2 \ominus r_1$$

$$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$$

$$\mathbb{B}(q,t) = (\mathit{true} \wedge \mathit{true}) \wedge \neg \mathit{true} = \mathit{false}$$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Free Structure Example

Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$$\mathbb{N}^{-}[I](q,t) = {r_1}^2 \ominus r_1$$

$$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$$

$$\mathbb{B}(q,t) = (true \wedge true) \wedge \neg true = false$$

$$eval_{\mathbb{N}}(\mathbb{N}^{-}[I](q,t)) = 2^2 - 2 = 2$$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Free Structure Example

Example

$$q = (R \bowtie R) - R \text{ with } r_1 = 2$$

$\mathbb{N}^{-}[I](q,t) = {r_1}^2 \ominus r_1$

$\mathbb{N}(q,t) = 2 \times 2 - 2 = 2$

$\mathbb{B}(q,t) = (true \wedge true) \wedge \neg true = false$

$eval_{\mathbb{N}}(\mathbb{N}^{-}[I](q,t)) = 2^2 - 2 = 2$

$eval_{\mathbb{B}}(\mathbb{N}^{-}[I](q,t)) = true \wedge true \wedge \neg true = false$

R

	a
r_1	1

Q_2

	a
${r_1}^2 \ominus r_1$	1

Outline

Lessons Learned

- Positive (monotonic) Operators
  - Provenance transitive
  - ⇒Approaches like Perm or WHIPS (Lineage) work well
- Negative Operators
  - Provenance is no longer transitive
  - minus minus is plus
  - Tuples not in provenance of any output tuple can influence result
  - ⇒Three approaches to deal with that
    - No right-hand side tuples in provenance
    - The broken Lineage approach
    - Model that goes beyond single operators

Relation between Set Difference and Negation

Definitions

Set difference contains all tuples from R where there exists no matching tuple from S

$$[[R - S]] = \{t \mid t \in R \wedge t \notin S\}$$

Equivalences (Set semantics)

$$R - S \equiv \sigma_{\neg \exists t \in \sigma_{\mathbf{R=S}}(S)}(R)$$

ILLINOIS INSTITUTE OF TECHNOLOGY

What about Outer Joins?

Outer Joins use Non-Existence of Tuples

$$[[R \ltimes\!\!\!\rtimes_C S]] = \{(t \blacktriangleright t') \mid t \in R \wedge t' \in S\} \cup \{(t_1 \blacktriangleright null(S)) \mid t \in R \wedge (\nexists t' \in S : (t \blacktriangleright t') \models C)\}$$

$\Rightarrow$

$$R \ltimes\!\!\!\rtimes_C S \equiv R \bowtie_C S \cup \pi_{\mathbf{R}, null(S)}(\sigma_{\neg \exists t \in \sigma_C(S)}(R))$$
$$R \ltimes\!\!\!\rtimes_C S \equiv R \bowtie_C S \cup \pi_{\mathbf{R}, null(S)}(R - (\pi_{\mathbf{R}}(R \bowtie_C S)))$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Set-difference and Aggregation

Set semantics

- Count the number of tuples in

$$
\begin{aligned}
R - S \equiv \pi_{\mathbf{R}}(& \\
& \alpha_{\mathbf{R},sum(x)}(\pi_{\mathbf{R},1 \rightarrow x}(R) \cup \pi_{\mathbf{S},1 \rightarrow x}(S)) \\
& \bowtie \alpha_{\mathbf{R},sum(x)}(\pi_{\mathbf{R},1 \rightarrow x}(R)) \\
&)
\end{aligned}
$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Outline

Recap

Provenance for Set-Difference

- Declarative models not sufficient (except Causality)
  - Insensitive Why + Perm Influence: no right-hand side tuples
  - Lineage: irrelevant right-hand side tuples
- Syntactic models
  - Undefined new operation $-$

ILLINOIS INSTITUTE OF TECHNOLOGY

Recap

Extending the Semiring Algebra

- $\mathbb{Z}$-relations fail
- Monus operation better, but
  - $\mathbb{N}[I]$ no longer free
  - Free object more complicated

Recap

Relation to other operators

- Set-difference as nested subqueries
- Left-outer join as set-difference
- Set-difference as aggregation

ILLINOIS INSTITUTE OF TECHNOLOGY

Literature

Todd J. Green, Zachary G. Ives, and Val Tannen.
Reconcilable differences.
In ICDT '09: Proceedings of the 16th International Conference on Database Theory, 212–224, Saint Petersburg, Russia, March 2009.

F. Geerts and A. Poggi.
On database query languages for K-relations.
Journal of Applied Logic, 8(2):173–185, 2010.

Yael Amsterdamer, Daniel Deutch, and Val Tannen.
On the limitations of provenance for queries with difference.
In TaPP '11: 3rd USENIX Workshop on the Theory and Practice of Provenance, 2011.

Y. Amsterdamer, D. Deutch, and V. Tannen.
Provenance for Aggregate Queries.
In PODS, 2011.

ILLINOIS INSTITUTE OF TECHNOLOGY