Fine-grained Provenance for Linear Algebra Operators

Zhepeng Yan    Val Tannen    Zachary Ives

University of Pennsylvania
Motivation

• Provenance is well-understood for relational data / queries.
  • E.g., view maintenance, delete propagation, computing trust, prob. db

• But increasingly analysts are performing more complex tasks:
  • Machine learning, data mining, image analysis, graph analytics

• Array data and matrix algebra are commonly used!
Motivation

• Provenance is well-understood for relational data / queries.
  • E.g., view maintenance, delete propagation, computing trust, prob. db

• But increasingly analysts are performing more complex tasks:
  • Machine learning, data mining, image analysis, graph analytics

• Array data and matrix algebra are commonly used!

• Question: How do we track provenance in this setting?
Inspiration: Provenance Semirings

• An algebra framework [Green et al. PODS’07] for
  • annotating tuples in a relation
  • propagating annotations through relational queries (SPJU and aggregation)

• Enables efficient delete propagation, view maintenance, etc
## Semiring example: input data

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID</td>
<td>PValue</td>
<td>QID</td>
</tr>
<tr>
<td>101</td>
<td>11</td>
<td>2003</td>
</tr>
<tr>
<td>102</td>
<td>12</td>
<td>2004</td>
</tr>
<tr>
<td>2003</td>
<td>13</td>
<td>2005</td>
</tr>
<tr>
<td>2004</td>
<td>14</td>
<td>2006</td>
</tr>
<tr>
<td>2005</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Semiring example: query

\[ S(\text{Value}) ::= P(x, \text{Value}), R(x, \_)
\]
\[ S(\text{Value}) ::= P(x, \text{Value}), Q(x, \_)
\]
\[ S(\text{Value}) ::= R(_, \text{Value})
\]
## Semiring example: query

### P

<table>
<thead>
<tr>
<th>PID</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>11</td>
</tr>
<tr>
<td>102</td>
<td>12</td>
</tr>
<tr>
<td>2003</td>
<td>13</td>
</tr>
<tr>
<td>2004</td>
<td>14</td>
</tr>
<tr>
<td>2005</td>
<td>15</td>
</tr>
<tr>
<td>2006</td>
<td>16</td>
</tr>
</tbody>
</table>

### Q

<table>
<thead>
<tr>
<th>QID</th>
<th>QValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>0</td>
</tr>
</tbody>
</table>

### R

<table>
<thead>
<tr>
<th>RID</th>
<th>RValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>13</td>
</tr>
<tr>
<td>102</td>
<td>14</td>
</tr>
<tr>
<td>5005</td>
<td>15</td>
</tr>
<tr>
<td>5006</td>
<td>16</td>
</tr>
<tr>
<td>5007</td>
<td>17</td>
</tr>
<tr>
<td>5008</td>
<td>18</td>
</tr>
<tr>
<td>5009</td>
<td>19</td>
</tr>
</tbody>
</table>

\[ S(\text{Value}) :: \text{P}(x, \text{Value}), \text{R}(x, \_). \]
\[ S(\text{Value}) :: \text{P}(x, \text{Value}), \text{Q}(x, \_). \]
\[ S(\text{Value}) :: \text{R}(\_, \text{Value}). \]
Semiring example: query

\[ S(\text{Value}) \triangleq P(x, \text{Value}), R(x, _) \]
\[ S(\text{Value}) \triangleq P(x, \text{Value}), Q(x, _) \]
\[ S(\text{Value}) \triangleq R(_, \text{Value}) \]
**Semiring example: query**

<table>
<thead>
<tr>
<th>PID</th>
<th>PValue</th>
<th>QID</th>
<th>QValue</th>
<th>RID</th>
<th>RValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>11</td>
<td>2003</td>
<td>0</td>
<td>101</td>
<td>13</td>
</tr>
<tr>
<td>102</td>
<td>12</td>
<td>2004</td>
<td>0</td>
<td>102</td>
<td>14</td>
</tr>
<tr>
<td>2003</td>
<td>13</td>
<td>2005</td>
<td>0</td>
<td>5005</td>
<td>15</td>
</tr>
<tr>
<td>2004</td>
<td>14</td>
<td>2006</td>
<td>0</td>
<td>5006</td>
<td>16</td>
</tr>
<tr>
<td>2005</td>
<td>15</td>
<td></td>
<td></td>
<td>5007</td>
<td>17</td>
</tr>
<tr>
<td>2006</td>
<td>16</td>
<td></td>
<td></td>
<td>5008</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5009</td>
<td>19</td>
</tr>
</tbody>
</table>

\[
S(\text{Value}) :: P(x, \text{Value}), R(x, \_)
\]

\[
S(\text{Value}) :: P(x, \text{Value}), Q(x, \_)
\]

\[
S(\text{Value}) :: R(_, \text{Value})
\]
Semiring example: output tuples

\[
\begin{array}{|c|}
\hline
\text{Value} \\
11 \\
12 \\
13 \\
14 \\
15 \\
16 \\
17 \\
18 \\
19 \\
\hline
\end{array}
\]

\[
S(\text{Value}) ::= \text{P}(x, \text{Value}), \text{R}(x, _) \\
S(\text{Value}) ::= \text{P}(x, \text{Value}), \text{Q}(x, _) \\
S(\text{Value}) ::= \text{R}(_, \text{Value})
\]
Semiring example: annotated output

\[
\begin{array}{|c|c|}
\hline
\text{Value} & \text{Annotation} \\
\hline
11 & pr \\
12 & pr \\
13 & pq + r \\
14 & pq + r \\
15 & pq + r \\
16 & pq + r \\
17 & r \\
18 & r \\
19 & r \\
\hline
\end{array}
\]

\[
S(V)  \quad S(V) \quad S(V) \quad S(V)
\]

\[
S(V) :: \ P(x, V) \quad R(x, _)
\]

\[
S(V) :: \ P(x, V) \quad Q(x, _)
\]

\[
S(V) :: \ R(_, V)
\]
Semiring example: delete propagation

What if we remove tuples from $P$?

Set $p = 0$!

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>$pr$</td>
</tr>
<tr>
<td>12</td>
<td>$pr$</td>
</tr>
<tr>
<td>13</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>14</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>15</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>16</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>17</td>
<td>$r$</td>
</tr>
<tr>
<td>18</td>
<td>$r$</td>
</tr>
<tr>
<td>19</td>
<td>$r$</td>
</tr>
</tbody>
</table>
Semiring example: delete propagation

What if we remove tuples from \( P \)?

Set \( p = 0! \)
Semimodule example: aggregation

Query: $\text{SUM}(\text{Value})$

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>$pr$</td>
</tr>
<tr>
<td>12</td>
<td>$pr$</td>
</tr>
<tr>
<td>13</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>14</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>15</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>16</td>
<td>$pq + r$</td>
</tr>
<tr>
<td>17</td>
<td>$r$</td>
</tr>
<tr>
<td>18</td>
<td>$r$</td>
</tr>
<tr>
<td>19</td>
<td>$r$</td>
</tr>
</tbody>
</table>
Semimodule example: annotation

Query: \( \text{SUM(Value)} \)

Annotated aggregation

\[
pr \times (11 + 12) + (pq + r) \times (13 + 14 + 15 + 16) + r \times (17 + 18 + 19)
\]
Semimodule example: annotation

Query: \( \text{SUM(Value)} \)

Annotated aggregation

\[
pr \cdot (11 + 12) + (pq + r) \cdot (13 + 14 + 15 + 16) + r \cdot (17 + 18 + 19)
\]

First term: annotation

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>( pr )</td>
</tr>
<tr>
<td>12</td>
<td>( pr )</td>
</tr>
<tr>
<td>13</td>
<td>( pq + r )</td>
</tr>
<tr>
<td>14</td>
<td>( pq + r )</td>
</tr>
<tr>
<td>15</td>
<td>( pq + r )</td>
</tr>
<tr>
<td>16</td>
<td>( pq + r )</td>
</tr>
<tr>
<td>17</td>
<td>( r )</td>
</tr>
<tr>
<td>18</td>
<td>( r )</td>
</tr>
<tr>
<td>19</td>
<td>( r )</td>
</tr>
</tbody>
</table>
Semimodule example: annotation

Query: SUM(Value)

Annotated aggregation

\[ pr \times (11 + 12) + (pq + r) \times (13 + 14 + 15 + 16) + r \times (17 + 18 + 19) \]

First term: annotation
Second term: value

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>pr</td>
</tr>
<tr>
<td>12</td>
<td>pr</td>
</tr>
<tr>
<td>13</td>
<td>pq + r</td>
</tr>
<tr>
<td>14</td>
<td>pq + r</td>
</tr>
<tr>
<td>15</td>
<td>pq + r</td>
</tr>
<tr>
<td>16</td>
<td>pq + r</td>
</tr>
<tr>
<td>17</td>
<td>r</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
</tr>
<tr>
<td>19</td>
<td>r</td>
</tr>
</tbody>
</table>
Semimodule example: annotation

S

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>pr</td>
</tr>
<tr>
<td>12</td>
<td>pr</td>
</tr>
<tr>
<td>13</td>
<td>pq + r</td>
</tr>
<tr>
<td>14</td>
<td>pq + r</td>
</tr>
<tr>
<td>15</td>
<td>pq + r</td>
</tr>
<tr>
<td>16</td>
<td>pq + r</td>
</tr>
<tr>
<td>17</td>
<td>r</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
</tr>
<tr>
<td>19</td>
<td>r</td>
</tr>
</tbody>
</table>

Query: SUM(Value)

Annotated aggregation

\[ pr \times (11 + 12) + (pq + r) \times (13 + 14 + 15 + 16) + r \times (17 + 18 + 19) \]

or \[ pr \times 23 + (pq + r) \times 58 + r \times 54 \]
Semimodule example: annotation

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>pr</td>
</tr>
<tr>
<td>12</td>
<td>pr</td>
</tr>
<tr>
<td>13</td>
<td>pq + r</td>
</tr>
<tr>
<td>14</td>
<td>pq + r</td>
</tr>
<tr>
<td>15</td>
<td>pq + r</td>
</tr>
<tr>
<td>16</td>
<td>pq + r</td>
</tr>
<tr>
<td>17</td>
<td>r</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
</tr>
<tr>
<td>19</td>
<td>r</td>
</tr>
</tbody>
</table>

Query: $\text{SUM(Value)}$

**Annotated aggregation**

$$pr \times (11 + 12) + (pq + r) \times (13 + 14 + 15 + 16) + r \times (17 + 18 + 19)$$

or

$$pr \times 23 + (pq + r) \times 58 + r \times 54$$

or

$$pq \times 58 + r \times (p \times 23 + 112)$$
Semimodule example: annotation

Query: SUM(Value)

<table>
<thead>
<tr>
<th>Value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>pr</td>
</tr>
<tr>
<td>12</td>
<td>pr</td>
</tr>
<tr>
<td>13</td>
<td>pq + r</td>
</tr>
<tr>
<td>14</td>
<td>pq + r</td>
</tr>
<tr>
<td>15</td>
<td>pq + r</td>
</tr>
<tr>
<td>16</td>
<td>pq + r</td>
</tr>
<tr>
<td>17</td>
<td>r</td>
</tr>
<tr>
<td>18</td>
<td>r</td>
</tr>
<tr>
<td>19</td>
<td>r</td>
</tr>
</tbody>
</table>

Annotated aggregation

\[ pr \times (11 + 12) + (pq + r) \times (13 + 14 + 15 + 16) \]
\[ + r \times (17 + 18 + 19) \]

or

\[ pr \times 23 + (pq + r) \times 58 + r \times 54 \]

or

\[ pq \times 58 + r \times (p \times 23 + 112) \]

Delete propagation:

set \( r = 0 \) and obtain \( pq \times 58 \)
Tracking matrix provenance

- We want to get the same benefits!
  - Algebraically manipulate annotated matrices
  - Hypothetical deletion
Tracking matrix data

$A$

$m$ samples

$n$ features
Tracking matrix data: partitioning

\[ A \]

\[ m \text{ samples} \]

\[ n \text{ features} \]
Tracking matrix data: annotating

$A$

$m$ samples

$n$ features
Tracking matrix data: annotating

$A$

$m$ samples

$n$ features
Tracking matrix data: annotating

\[ A \]

\[ m \text{ samples} \]

\[ n \text{ features} \]
Tracking matrix data: annotating

\[ A = \begin{bmatrix} xu \end{bmatrix} \]

- \( m \) samples
- \( n \) features
Tracking matrix data: annotating

\[ A \]

\[ m \text{ samples} \]

\[ n \text{ features} \]
Challenges

• Specify and relate different parts of a given matrix
  • Matrix decomposition through selector matrices

• Specify connection between derived and source matrices
  • Embed matrix algebra and provenance into a semialgebra
Decomposition

\[ A = \begin{bmatrix} u & v \\ B & C \\ D & E \end{bmatrix} \]
Decomposition: selectors

\[ A = S_x B \ T_u + S_x C \ T_v + S_y D \ T_u + S_y E \ T_v \]

with selectors \( S_x, S_y, T_u, T_v \)
Decomposition: selectors

\[ A = S_x B T_u + S_x C T_v + S_y D T_u + S_y E T_v \]

with selectors \( S_x S_y T_u T_v \)
Decomposition: selectors

\[ A = S_x B T_u + S_x C T_v + S_y D T_u + S_y E T_v \]

with selectors \( S_x S_y T_u T_v \)
Decomposition: selectors

\[ A = S_x B T_u + S_x C T_v + S_y D T_u + S_y E T_v \]

with selectors \( S_x \) \( S_y \) \( T_u \) \( T_v \)
Decomposition: selectors

\[ A = S_x B T_u + S_x C T_v + S_y D T_u + S_y E T_v \]

with selectors \( S_x \) \( S_y \) \( T_u \) \( T_v \)
Summary: selectors

- Relate a matrix to its sub-matrices.
- Matrices with only 0/1.
- Any row / column has at most a 1.
Summary: selectors

- Relate a matrix to its sub-matrices.
- Matrices with only 0/1.
- Any row / column has at most a 1.

- Extends to non-adjacent case.
- Works for any rectangular partition.
Provenance propagation

• We have
  • Matrices and operators over them – Algebra of matrices $\mathcal{M}$
  • Annotations – Semiring of provenance polynomials $\mathbb{N}[X]$  

• Goals
  • Combine annotations in the same structure as the matrices
  • Operations should propagate data and annotations
Provenance propagation

• We have
  • Matrices and operators over them – Algebra of matrices $\mathcal{M}$
  • Annotations – Semiring of provenance polynomials $\mathbb{N}[X]$

• Goals
  • Combine annotations in the same structure as the matrices
  • Operations should propagate data and annotations

• We do this in the space of tensor product $\mathbb{N}[X] \otimes \mathcal{M}$
  • Matrices as vectors, provenance as scalars: $p \ast A$
  • Satisfies all the laws of a $\mathbb{N}[X]$-semialgebra.
Laws of a $\mathbb{N}[X]$-semialgebra ($K$-semialgebra)

$$(K, +_K, \cdot_K, 0_K, 1_K)$$ commutative semiring

$$(K \otimes \mathcal{M}, +, \cdot, 0, 1)$$ forms a ring (just like the matrices)

laws for scalar multiplication in a $K$-semialgebra

\[
\begin{align*}
    k \ast (A_1 + A_2) &= k \ast A_1 + k \ast A_2 \\
    k \ast 0 &= 0 \\
    (k_1 +_K k_2) \ast A &= k_1 \ast A + k_2 \ast A \\
    0_K \ast A &= 0 \\
    (k_1 \cdot_K k_2) \ast A &= k_1 \ast (k_2 \ast A) \\
    1_K \ast A &= A \\
    (k_1 \ast A_1)(k_2 \ast A_2) &= (k_1 \cdot_K k_2) \ast (A_1 A_2)
\end{align*}
\]
Semialgebra example

\[ A = S_x B T_u + S_x C T_v + S_y D T_u + S_y E T_v \]
Semialgebra example: add annotation

\[ A = S_x xu \cdot B T_u + S_x xv \cdot C T_v + S_y yu \cdot D T_u + S_y vy \cdot E T_v \]
Semialgebra example: propagate annotation

\[ A = S_x x_u B T_u + S_x x_v C T_v + S_y y_u D T_u + S_y v_y E T_v \]
Propagating annotation: transposition

\[ A = S_x xu^*B T_u + S_x xv^*C T_v \]
\[ + S_y yu^*D T_u + S_y vy^*E T_v \]

\[ A^T = T_{u}^T (xu \ast B^T) S_x^T + T_{u}^T (yu \ast D^T) S_y^T \]
\[ + T_{v}^T (xv \ast C^T) S_x^T + T_{v}^T (yv \ast E^T) S_y^T \]
Propagating annotation: transposition

\[ A = S_x x u \ast B \ T_u + S_x x v \ast C \ T_v \]
\[ + S_y y u \ast D \ T_u + S_y y v \ast E \ T_v \]

\[
A^T = T_u^T (x u \ast B^T) S_x^T + T_u^T (y u \ast D^T) S_y^T \\
+ T_v^T (x v \ast C^T) S_x^T + T_v^T (y v \ast E^T) S_y^T
\]

Transposition of a selector is still a selector
Still a sum of \((\text{selector} \times \text{matrix} \times \text{selector})\)
Propagating annotation: multiplication

\[ A = S_x \ x_u B \ T_u + S_x \ x_v C \ T_v + S_y \ y_u D \ T_u + S_y \ y_v E \ T_v \]

\[ A^T = T_u^T \ (x_u * B^T) \ S_x^T + T_u^T \ (y_u * D^T) \ S_y^T + T_v^T \ (x_v * C^T) \ S_x^T + T_v^T \ (y_v * E^T) \ S_y^T \]
Propagating annotation: multiplication

\[ A = S_x xu * B T_u + S_x xv * C T_v + S_y yu * D T_u + S_y vy * E T_v \]

\[ A^T = T_u^T (xu * B^T) S_x^T + T_u^T (yu * D^T) S_y^T + T_v^T (xv * C^T) S_x^T + T_v^T (vy * E^T) S_y^T \]

\[ A A^T = S_x (x^2 u^2 * B B^T + x^2 v^2 * C C^T) S_x^T + S_x (xy u^2 * B D^T + xy v^2 * C E^T) S_y^T + S_y (xy u^2 * D B^T + xy v^2 * E C^T) S_x^T + S_y (y^2 u^2 * D D^T + y^2 v^2 * E E^T) S_y^T \]
Propagating annotation: multiplication

\[
A = S_x xu^*B_T u + S_x xv^*C_T v + S_y yu^*D_T u + S_y vy^*E_T v
\]

\[
A^T = T_u^T (xu \ast B^T) S_x^T + T_u^T (yu \ast D^T) S_y^T + T_v^T (xv \ast C^T) S_x^T + T_v^T (yv \ast E^T) S_y^T
\]

\[
AA^T = S_x (x^2u^2 \ast BB^T + x^2v^2 \ast CC^T) S_x^T + S_x (xyu^2 \ast BD^T + xyv^2 \ast CE^T) S_y^T + S_y (xyu^2 \ast DB^T + xyv^2 \ast EC^T) S_x^T + S_y (y^2u^2 \ast DD^T + y^2v^2 \ast EE^T) S_y^T
\]

\[
T_u T_u^T = T_v T_v^T = 1
\]

\[
T_u T_v^T = T_v T_u^T = 0
\]
Semialgebra: delete propagation

\[ AA^T = S_x(x^2u^2 \ast BB^T + x^2v^2 \ast CC^T)S_x^T \]
\[ + S_x(xyu^2 \ast BD^T + xyv^2 \ast CE^T)S_y^T \]
\[ + S_y(xyu^2 \ast DB^T + xyv^2 \ast EC^T)S_x^T \]
\[ + S_y(y^2u^2 \ast DD^T + y^2v^2 \ast EE^T)S_y^T \]
Semialgebra: delete propagation

\[ AA^T = S_x (x^2 u^2 * BB^T + x^2 v^2 * CC^T) S_x^T + S_x (xyu^2 * BD^T + xyv^2 * CE^T) S_y^T + S_y (xyu^2 * DB^T + xyv^2 * EC^T) S_x^T + S_y (y^2 u^2 * DD^T + y^2 v^2 * EE^T) S_y^T \]

Deletion propagation: set \( y = 0 \)

\[ AA^T = x^2 * (u^2 * S_x BB^T S_x^T + v^2 * S_x CC^T S_x^T) \]
Preliminary application: solving equations

- $(A + B) x = b$, $A$ and $B$ are square matrices
- $A$ is from source $p$, $B$ is from source $q$
Preliminary application: solving equations

- $(A + B)x = b$, $A$ and $B$ are square matrices
- $A$ is from source $p$, $B$ is from source $q$

- Jacobi method: iteratively compute

$$u_{k+1} = (M^{-1}N)u_k + M^{-1}b$$
$$u_0 = \bar{0}$$

- $M = p \times \text{diag}(A)$, $N = p \times (\text{diag}(A) - A) - q \times B$
Jacobi method: example

• Iteratively compute
  \[ u_{k+1} = (M^{-1}N)u_k + M^{-1}b \quad u_0 = \vec{0} \]

• \( M = p \ast \text{diag}(A) \), \( N = p \ast (\text{diag}(A) - A) - q \ast B \)
Jacobi method: example

- Iteratively compute
  \[ u_{k+1} = (M^{-1}N)u_k + M^{-1}b \quad u_0 = \bar{0} \]
- \( M = p \ast \text{diag} (A) \), \( N = p \ast (\text{diag}(A) - A) - q \ast B \)

\[
A = p \ast \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad B = q \ast \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\]

\[
u_1 = p \ast \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \quad u_2 = p \ast \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} + p^3 \ast \begin{bmatrix} \frac{1}{4} \\ -\frac{1}{4} \end{bmatrix} + p^2 q \ast \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}, \ldots
\]
Preliminary applications

• Solving systems of linear equations

• Also in the paper
  • Largest eigenvalue
  • PageRank
Contributions

• First steps towards a semantics-preserving notion of fine-grained provenance for linear algebra operators
  • Key development: decomposition, tensor-product construction, and algebraic rules

• Preliminary applications in solving equations, computing largest eigenvalues, and PageRank.

• Key benefit
  • Automatic propagation of annotations through operators
  • Ability to assign values (e.g., 0 or 1) to the annotations and propagate the effects, e.g., for deletion or trust
Related and future work

- Provenance Semirings / Semimodules
  - Green et al. PODS’07, Amsterdamer et al. PODS’11

- Array databases
  - SciDB, RasDaMan
  - Wu et al. SubZero, Peng and Diao SIGMOD’15

- Distributed machine learning / linear algebra programs
  - SystemML, Spark, MLbase, Cumulon, MADlib, GraphX, LINView, etc

- Future work
  - Support more linear algebra operators
  - Scalable implementation
Thank you!