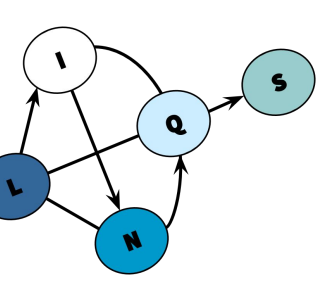# A Comparison of Bottom-Up Approaches to Grounding for Templated Markov Random Fields

Eriq Augustine & Lise Getoor
University of California, Santa Cruz

Grounding, the process of instantiating a templated Markov Random Field (MRF), is typically the limiting factor when modeling with MRFs. Here we investigate **three** key components in a MRF grounding system:

1. Inference Targets & Scoping
2. Trivial Potential Removal
3. Blocking

Grounding Templated Markov Random Fields

Rules

```
WritesAbout(A1, T) & WritesAbout(A2, T)
    => SameAuthor(A1, A2)
SimilarName(A1, A2) => SameAuthor(A1, A2)
SameAuthor(A1, A2) & SameAuthor(A2, A3)
    => SameAuthor(A1, A3)
```
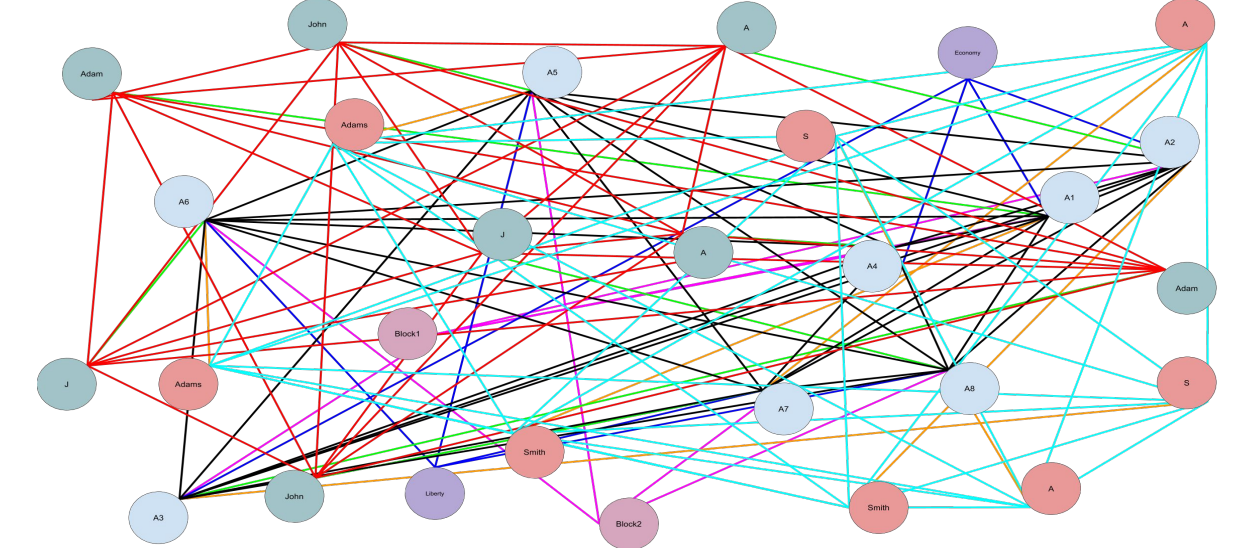
Data

WritesAbout

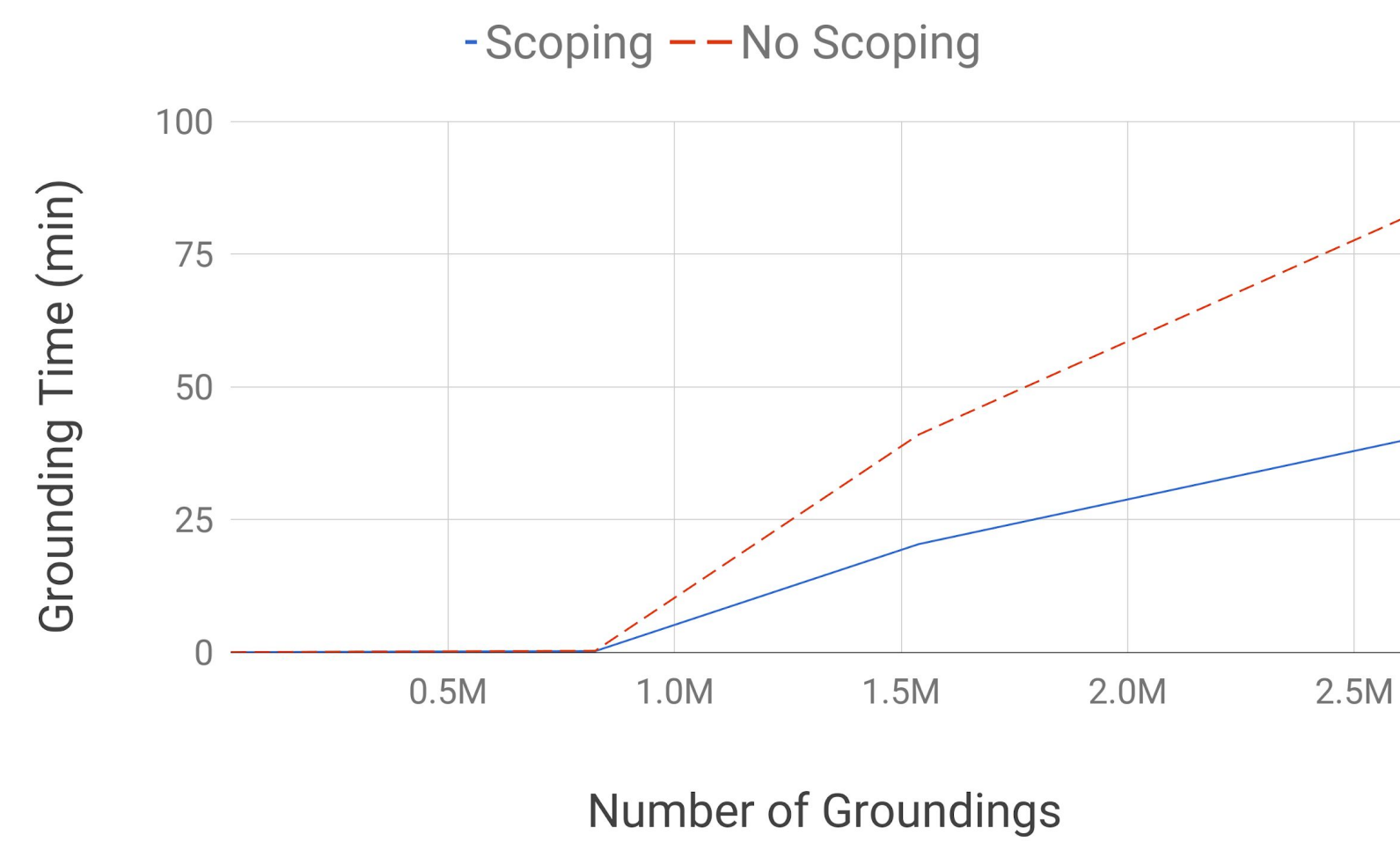|	
A Smith	Economy
Adam Smith	Economy
John Adams	Liberty
J Adams	Liberty

Grounding

Ground Rules

```
WritesAbout('A Smith', 'Economy') & WritesAbout('Adam Smith', 'Economy')
    => SameAuthor('A Smith', 'Adam Smith')
WritesAbout('Adam Smith', 'Economy') & WritesAbout('A Smith', 'Economy')
    => SameAuthor('Adam Smith', 'A Smith')
SameAuthor('A Smith', 'Adam Smith') & SameAuthor('Adam Smith', 'J Adams')
    => SameAuthor('A Smith', 'J Adams')
SameAuthor('John Adams', 'Adam Smith') & SameAuthor('Adam Smith', 'J Adams')
    => SameAuthor('John Adams', 'J Adams')
...
```

Ground Markov Random Field

1. Inference Targets & Scoping

```
WritesAbout(A1, T) & WritesAbout(A2, T)
    => SameAuthor(A1, A2)
```

- Scoping --- No Scoping

Grounding Time (min) vs Number of Groundings

Explicit Targets

SameAuthor

A Smith	Adam Smith
A Smith	J Adams
A Smith	John Adams
Adam Smith	J Adams
Adam Smith	John Adams
J Adams	John Adams

Implicit Targets

WritesAbout

A Smith	Economy
Adam Smith	Economy
John Adams	Liberty
J Adams	Liberty

SameAuthor

A Smith	A Smith
A Smith	Adam Smith
A Smith	John Adams
A Smith	J Adams
Adam Smith	A Smith
Adam Smith	Adam Smith
Adam Smith	John Adams
Adam Smith	J Adams
John Adams	A Smith
John Adams	Adam Smith
John Adams	John Adams
John Adams	J Adams
J Adams	A Smith
J Adams	Adam Smith
J Adams	John Adams
J Adams	J Adams

Build Type Domain

Authors

A Smith
Adam Smith
John Adams
J Adams

Topics

Economy
Liberty

Cross Product
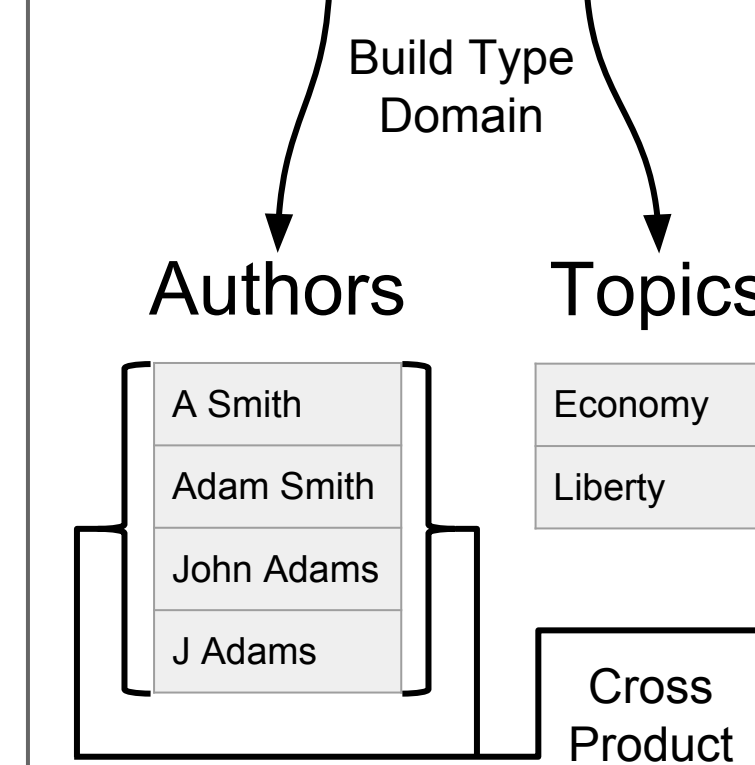
Implicit Targets With Scoping

```
SameAuthor(A1, A2) :=
SameFirstInitial(A1, A2) & (A1 < A2)
```

SameFirstInitial

A Smith	Adam Smith
Adam Smith	A Smith
John Adams	J Adams
J Adams	John Adams

Evaluate A1 < A2
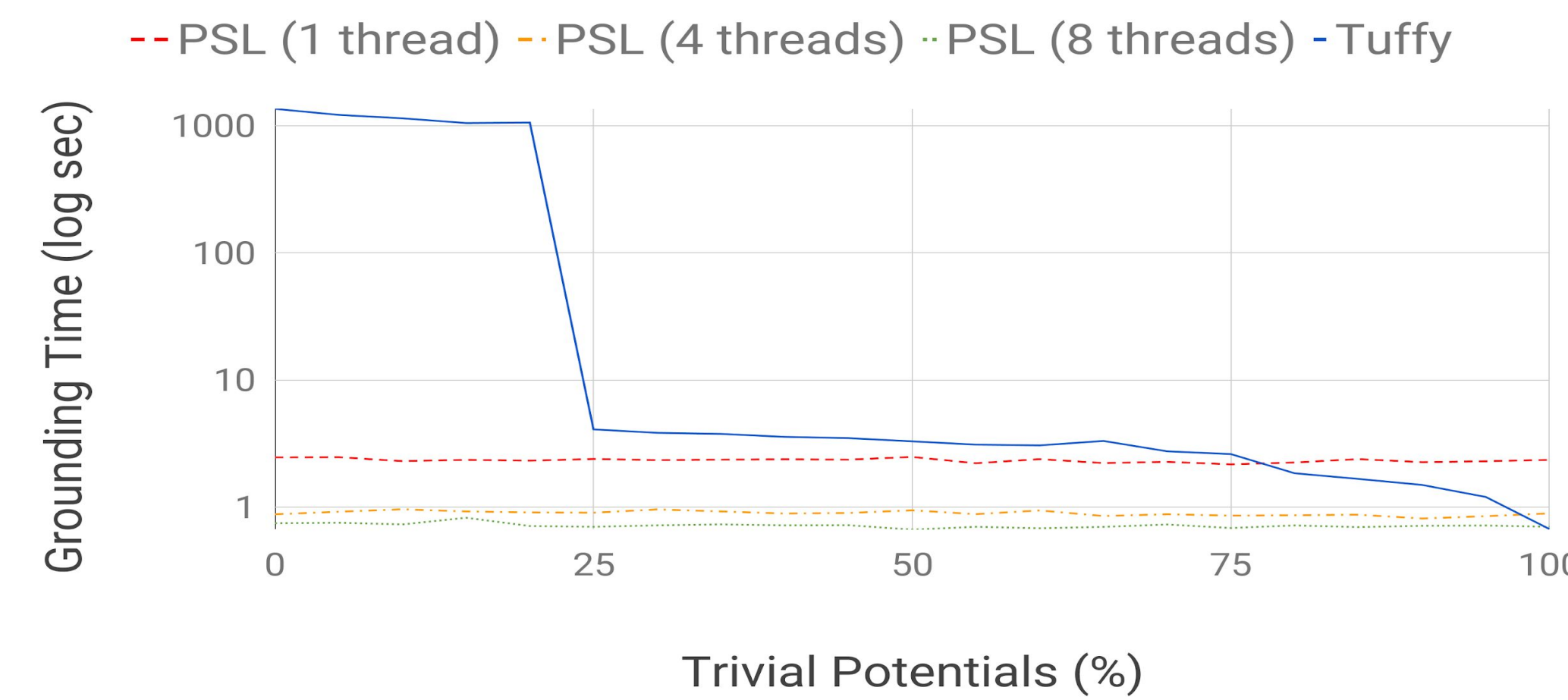
SameAuthor

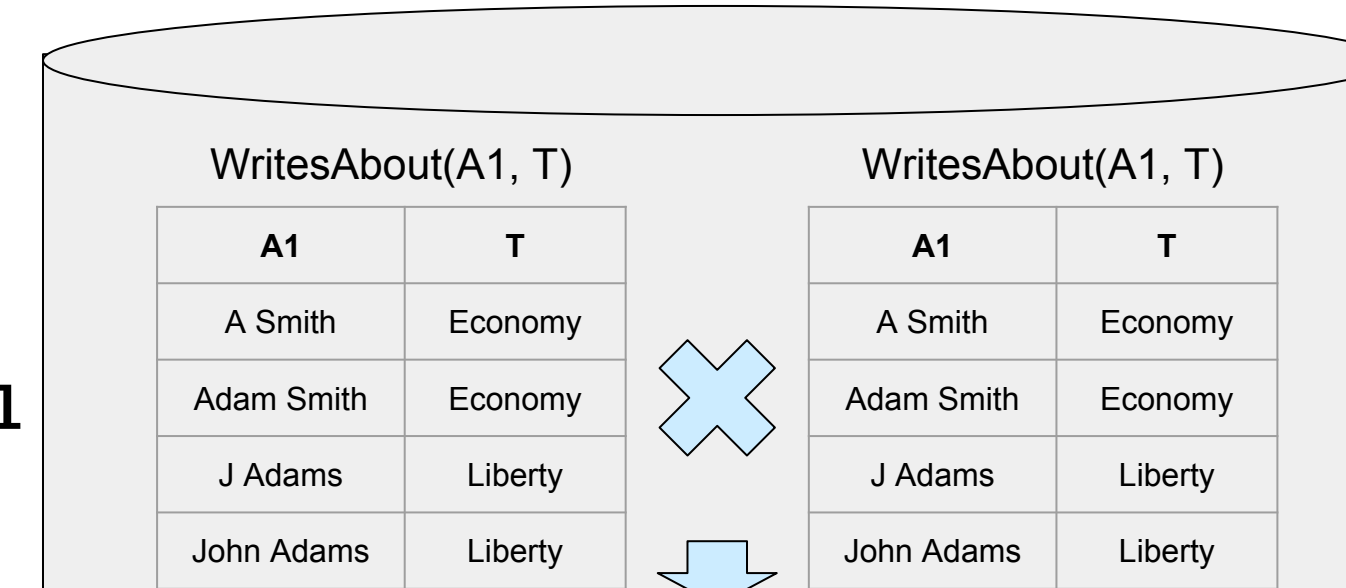|	
A Smith	Adam Smith
J Adams	John Adams

2. Trivial Potential Removal

```
WritesAbout(A1, T) & WritesAbout(A2, T) => SameAuthor(A1, A2)
```

```
!WritesAbout(A1, T) v !WritesAbout(A2, T) v SameAuthor(A1, A2)
```

False v **False** v **True** = **Trivial**

-- PSL (1 thread) ·· PSL (4 threads) ·· PSL (8 threads) - Tuffy

Grounding Time (log sec) vs Trivial Potentials (%)

Database Removal

WritesAbout(A1, T)

A1	T
A Smith	Economy
Adam Smith	Economy
J Adams	Liberty
John Adams	Liberty

WritesAbout(A1, T)

A1	T
A Smith	Economy
Adam Smith	Economy
J Adams	Liberty
John Adams	Liberty

```
A1 = 'A Smith', A2 = 'Adam Smith', T = 'Economy'
A1 = 'A Smith', A2 = 'John Adams', T = 'Economy'
A1 = 'J Adams', A2 = 'John Adams', T = 'Liberty'
A1 = 'J Adams', A2 = 'Adam Smith', T = 'Liberty'
```
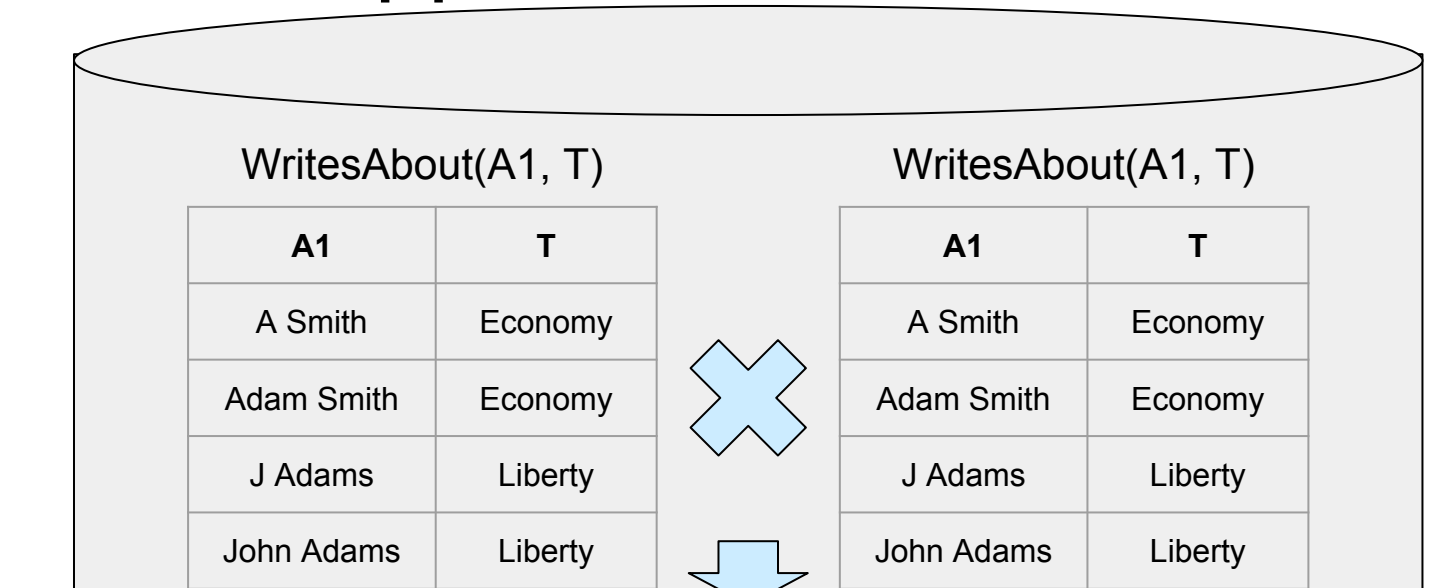
```
WritesAbout('A Smith', 'Economy') & WritesAbout('Adam Smith', 'Economy')
    => SameAuthor('A Smith', 'Adam Smith')
WritesAbout('A Smith', 'Economy') & WritesAbout('John Adams', 'Economy')
    => SameAuthor('A Smith', 'John Adams')
WritesAbout('J Adams', 'Liberty') & WritesAbout('John Adams', 'Liberty')
    => SameAuthor('J Adams', 'John Adams')
WritesAbout('J Adams', 'Liberty') & WritesAbout('Adam Smith', 'Liberty')
    => SameAuthor('J Adams', 'Adam Smith')
```
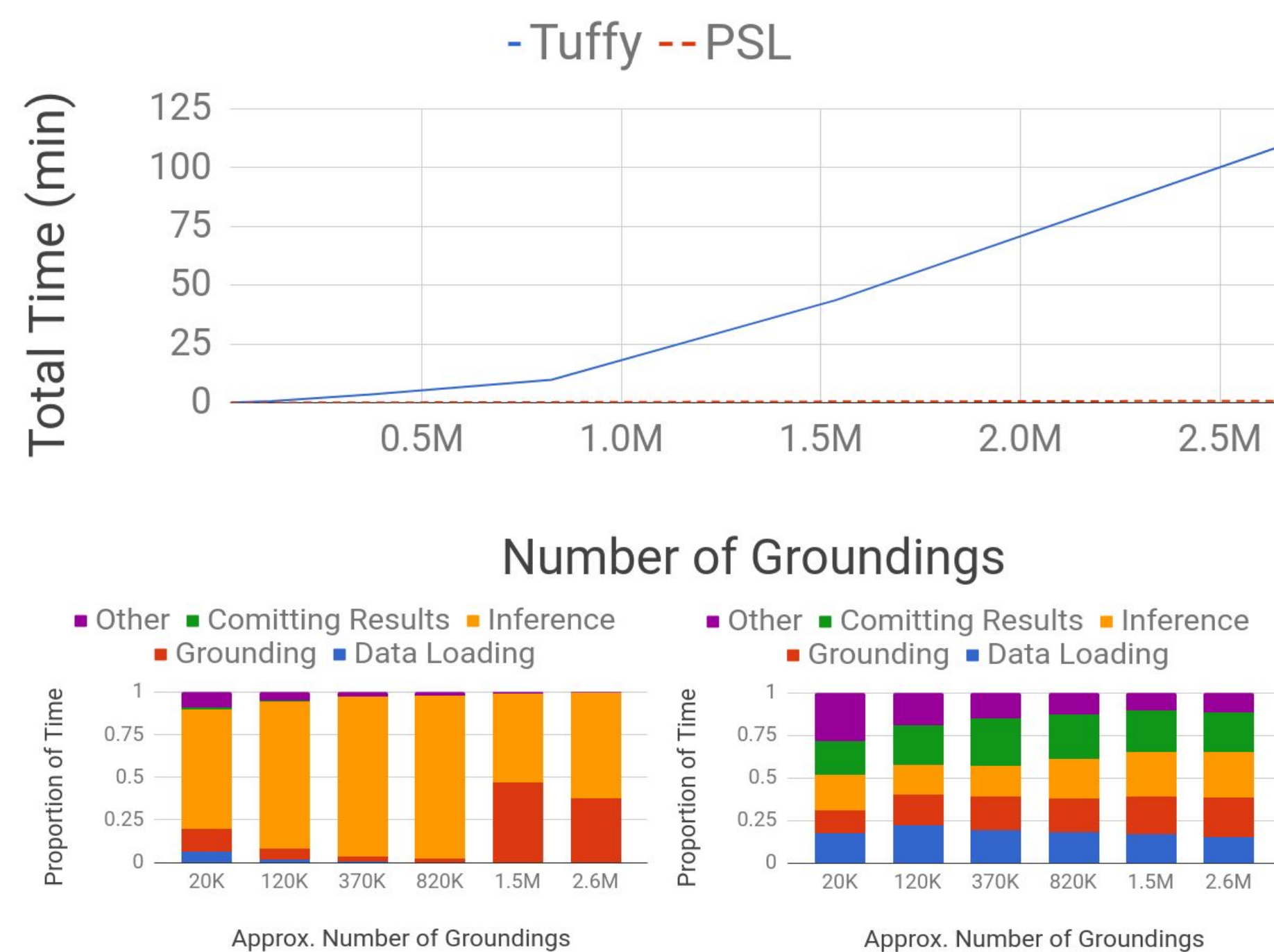
```
WritesAbout('A Smith', 'Economy') & WritesAbout('Adam Smith', 'Economy')
    => SameAuthor('A Smith', 'Adam Smith')
WritesAbout('J Adams', 'Liberty') & WritesAbout('John Adams', 'Liberty')
    => SameAuthor('J Adams', 'John Adams')
```

Application Removal

WritesAbout(A1, T)

A1	T
A Smith	Economy
Adam Smith	Economy
J Adams	Liberty
John Adams	Liberty

WritesAbout(A1, T)

A1	T
A Smith	Economy
Adam Smith	Economy
J Adams	Liberty
John Adams	Liberty

```
A1 = 'A Smith', A2 = 'Adam Smith', T = 'Economy'
A1 = 'A Smith', A2 = 'John Adams', T = 'Economy'
A1 = 'J Adams', A2 = 'John Adams', T = 'Liberty'
A1 = 'J Adams', A2 = 'Adam Smith', T = 'Liberty'
```

```
WritesAbout('A Smith', 'Economy') & WritesAbout('Adam Smith', 'Economy')
    => SameAuthor('A Smith', 'Adam Smith')
WritesAbout('J Adams', 'Liberty') & WritesAbout('John Adams', 'Liberty')
    => SameAuthor('J Adams', 'John Adams')
```
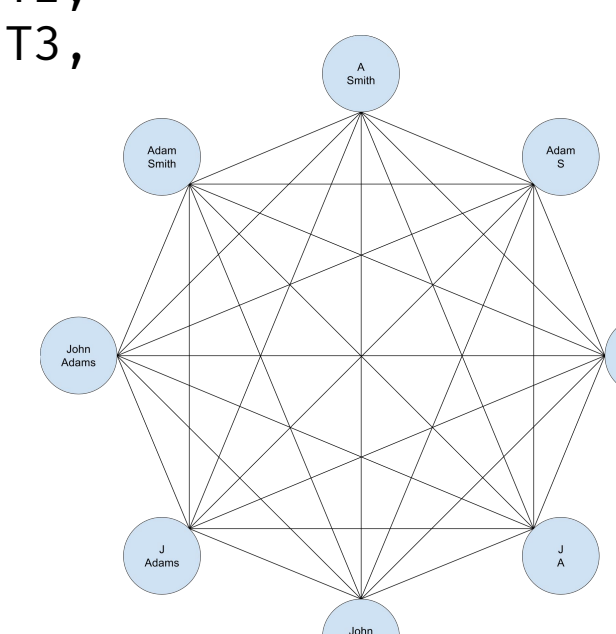
End to End Comparison

- Tuffy -- PSL

Total Time (min) vs Number of Groundings

Other, Comitting Results, Inference, Grounding, Data Loading

Proportion of Time vs Approx. Number of Groundings (20K, 120K, 370K, 820K, 1.5M, 2.6M)

3. Blocking

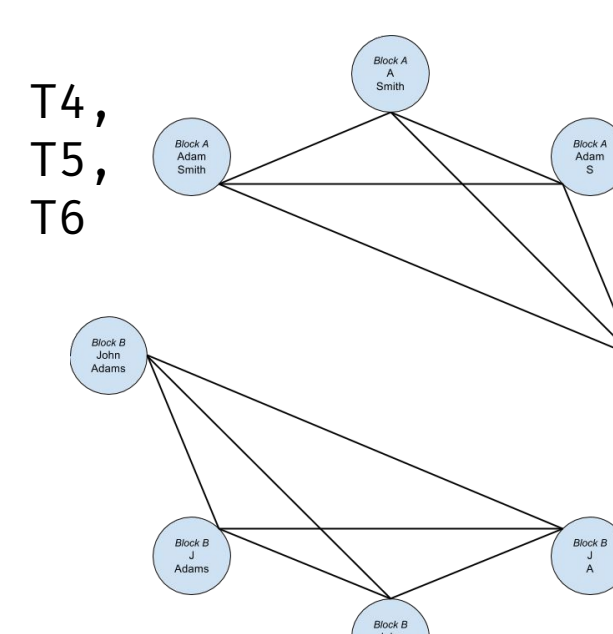|	No Blocking	Implicit Blocking	Explicit Blocking
Rule	`SameAuthor(A1, A2) & SameAuthor(A2, A3) => SameAuthor(A1, A3)`	`Block(A1, X) & Block(A2, X) & Block(A3, X) & SameAuthor(A1, A2) & SameAuthor(A2, A3) => SameAuthor(A1, A3)`	`Block(A1, X) & Block(A2, X) & Block(A3, X) & SameAuthor(A1, A2) & SameAuthor(A2, A3) => SameAuthor(A1, A3)`
Query	`SELECT ... FROM SameAuthor T1, SameAuthor T2, SameAuthor T3, WERE ...`	`SELECT ... FROM Block T1, Block T2, Block T3, SameAuthor T4, SameAuthor T5, SameAuthor T6 WERE ...`	`SELECT ... FROM Block T1, Block T2, Block T3 WERE ...`
Result Graph			

Query Time (log ms) — Similarity, Symmetry, Transitivity

Query Result Size — Similarity, Symmetry, Transitivity

Max Query Node Size — Similarity, Symmetry, Transitivity