A1 Presentation

Ben Athiwaratkun and Keegan Kang

September 15, 2014

Presentation Objectives

Introduction

- Most NLP goals focus on making some kind of prediction about a given text.

Introduction

- Most NLP goals focus on making some kind of prediction about a given text.
- Our proposal is to see if we can make some kind of prediction about the *people* who write these texts (social interaction part of the course!).

Research Problem

Can we predict the dynamics of any two active users, based on their previous posts?

Research Problem

Can we predict the dynamics of any two active users, based on their previous posts?

We use the Slashdog blog conversations from BC3's blog corpus in our investigation.

Research Problem (in detail)

1. We would like to know if we can predict the trend of a thread, given the earliest k comments in the article.

Research Problem (in detail)

1. We would like to know if we can predict the trend of a thread, given the earliest k comments in the article.
2. Given the probability distribution for any user A to create posts of quality, we would like to estimate the conditional probability vector of user A, given that some event $\mathcal{E}$ has occurred.

Proposed techniques to tackle the problem

1. Look at frequencies!

Proposed techniques to tackle the problem

1. Look at frequencies!
2. Plot the comment length in each thread to see if there are any discernible patterns.

Proposed techniques to tackle the problem

1. Look at frequencies!
2. Plot the comment length in each thread to see if there are any discernible patterns.
3. Compare distributions and conditional distributions of several metrics.

Example of metrics

1. [Ott09]

Example of metrics

1. [Ott09]
2. "Connectedness", as in [BKLDNM13], by looking at time intervals

Classification Breakdown

A quick breakdown by category...

Classification Breakdown

A quick breakdown by category...

```
{'Troll': 14932, 'Funny': 40672, 'None': 464104,
'Flamebait': 7456,  'Redundant': 4792, 'Offtopic': 11384,
'Informativ': 40188, 'Interestin':  50168,
'Insightful': 73864}
```

Classification Breakdown

...and by dominating class (when None is removed)

Classification Breakdown

...and by dominating class (when None is removed)

```
{'Flamebait': 28, 'Funny': 2316, 'Redundant': 8,
'Troll': 136,  'Offtopic': 68, 'Insightful': 5540,
'Interestin': 1824, 'Informativ': 1348}
```

Classification Breakdown

...and by dominating class (when None is removed)

```
{'Flamebait': 28, 'Funny': 2316, 'Redundant': 8,
'Troll': 136,  'Offtopic': 68, 'Insightful': 5540,
'Interestin': 1824, 'Informativ': 1348}
```

We could potentially try to classify the comments with class `None` with some the tools Python provides.

Example of Comment Lengths

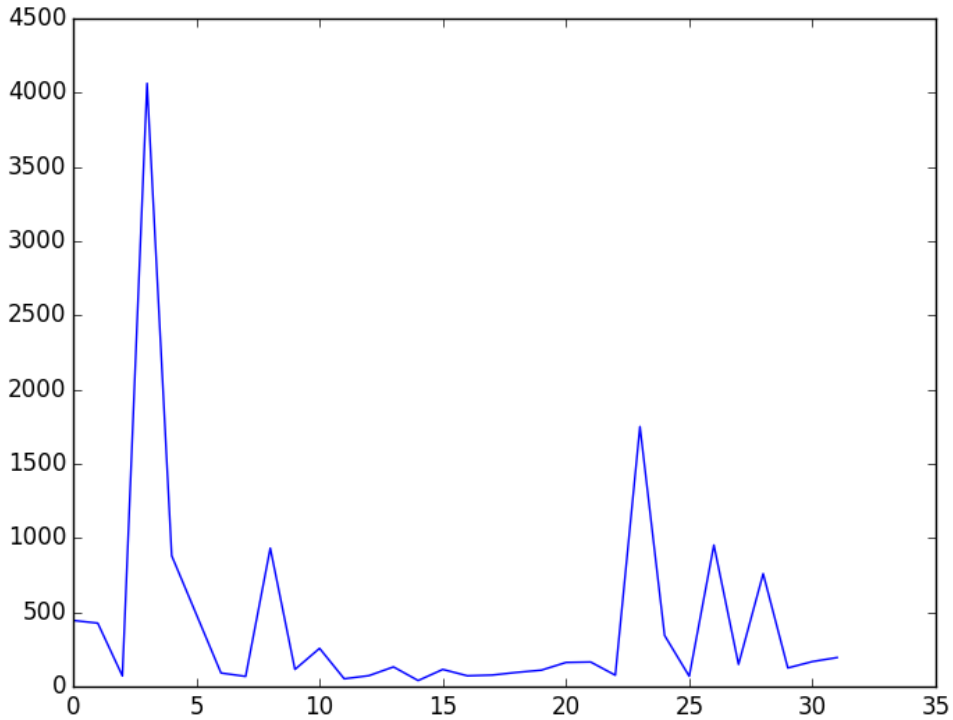


Figure : Post Length for Article `Conference Board Admits Plagiarism, Pulls Copyright Report`

Example of Comment Lengths

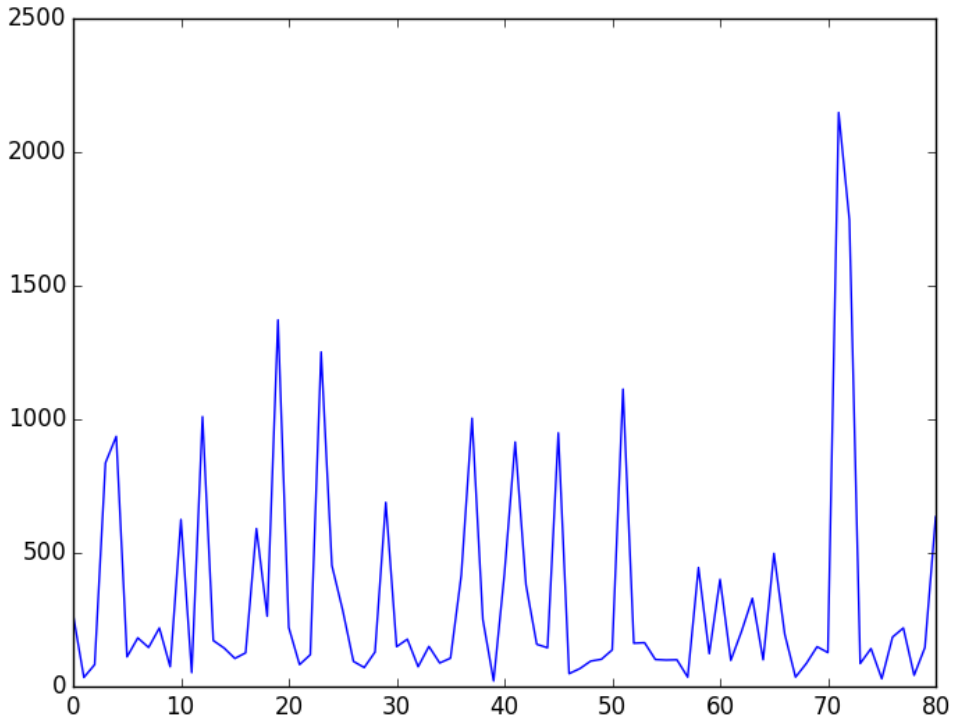


Figure : Post Length for Article Google's "Wave" Blurs Chat, Email, Collaboration Software

Distribution and Conditional Distribution of Post Lengths

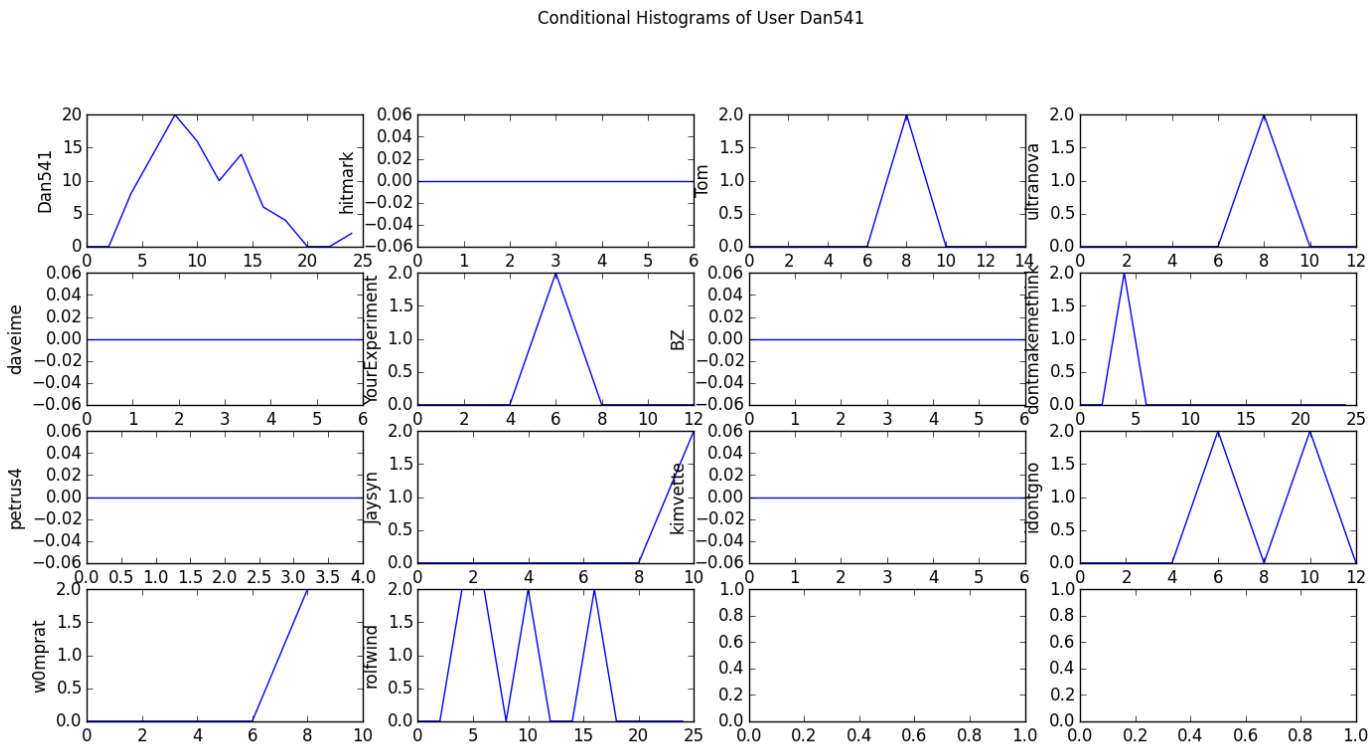


Figure : Joint Distribution Plots for user `Dan541` and Other Active User

Distribution and Conditional Distribution of Post Lengths

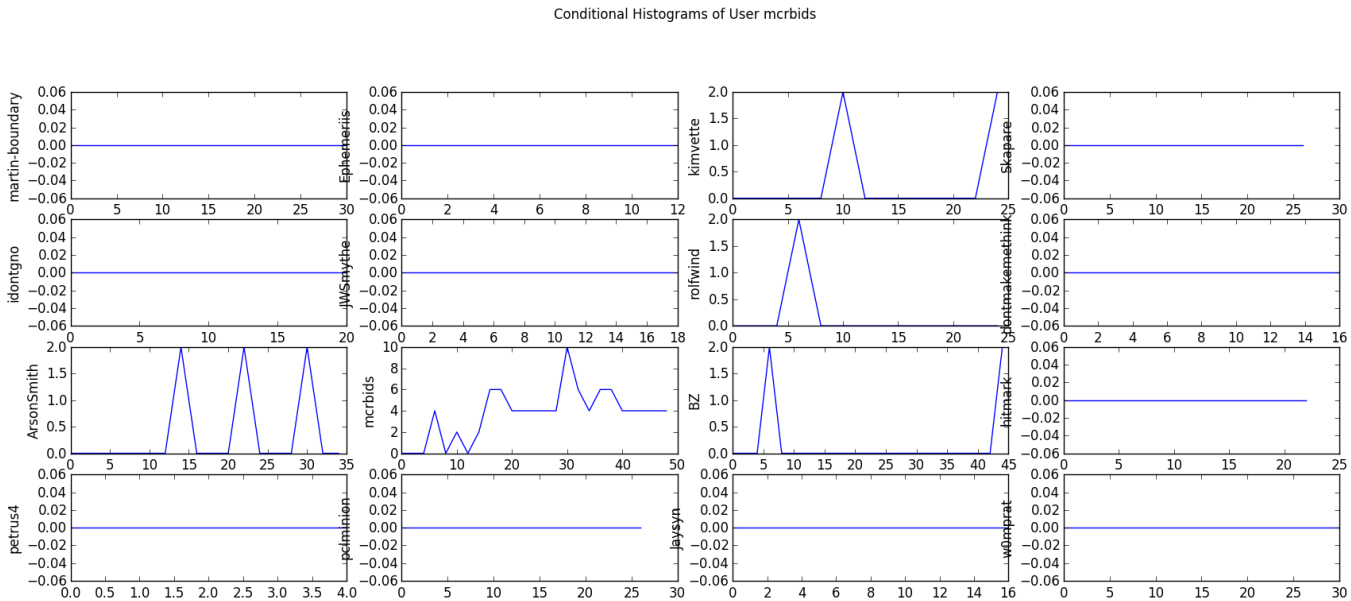
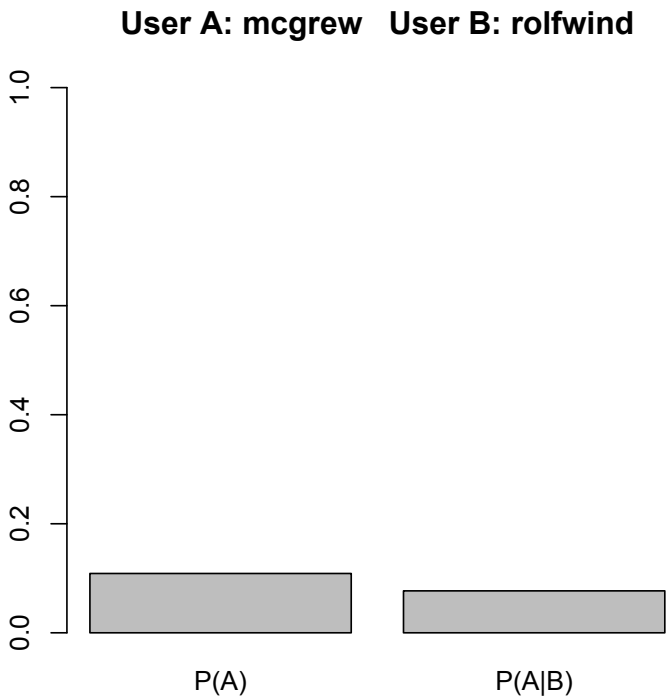
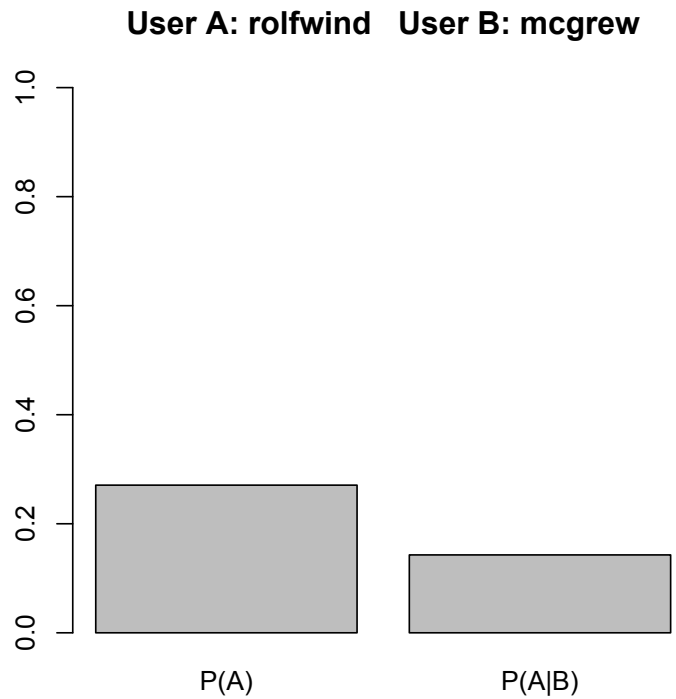


Figure : Joint Distribution Plots for user `mcrbids` and Other Active User

Probability and Conditional Probability of Post Type

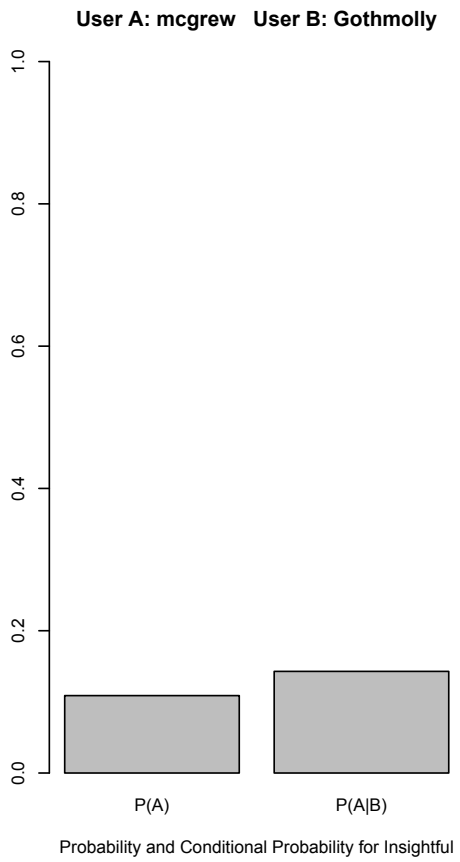


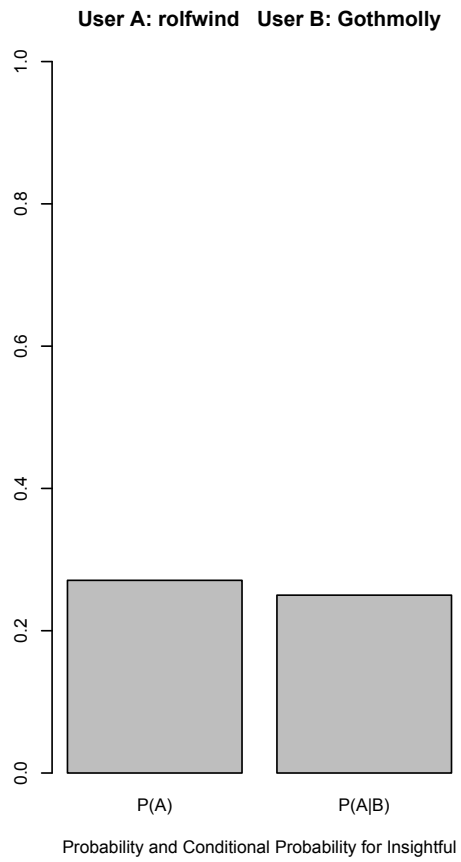
Probability and Conditional Probability for Insightful

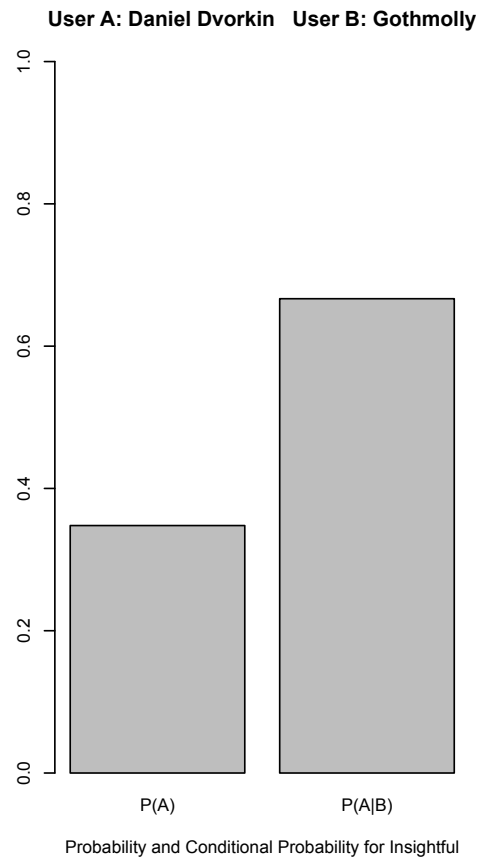


Probability and Conditional Probability for Insightful

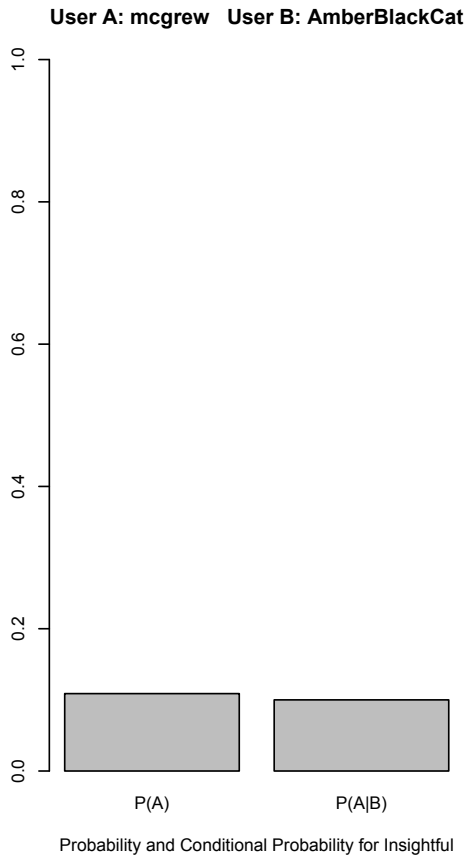
Probability and Conditional Probability of Post Type

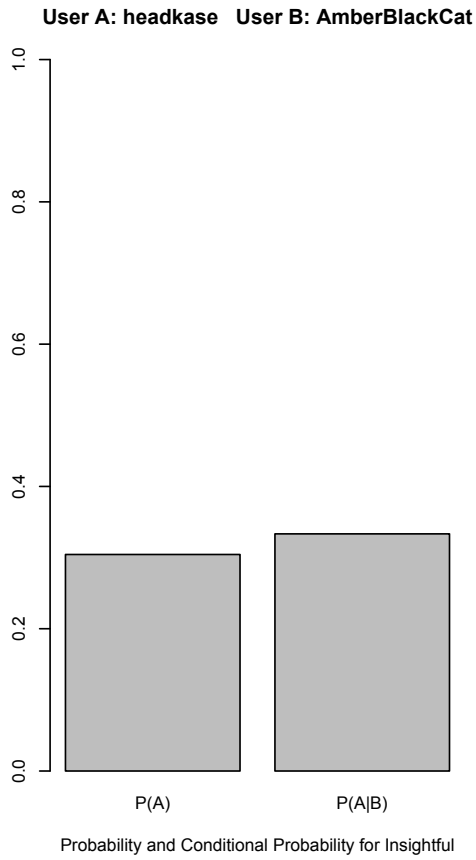


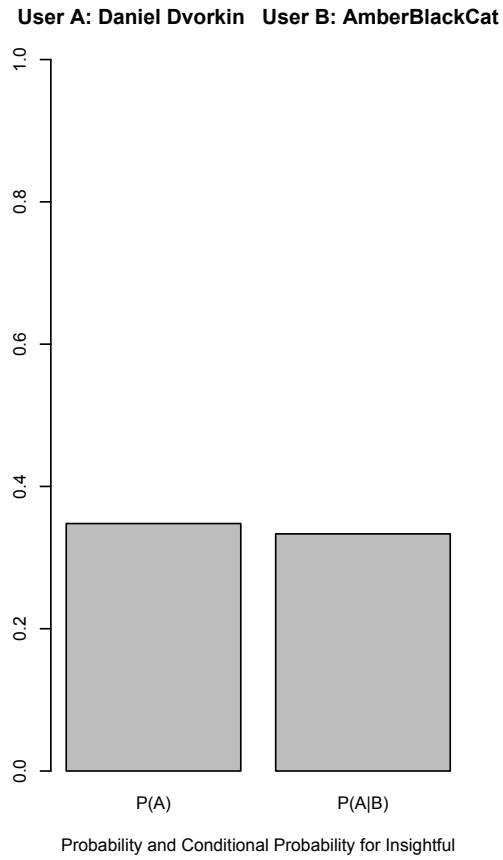




Probability and Conditional Probability of Post Type







Discussion

1. We want to differentiate the real signal from noise.

Discussion

1. We want to differentiate the real signal from noise.
2. How can we infer social interaction from a given signal?


Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil.
Characterizing and curating conversation threads: Expansion, focus, volume, re-entry.
In *Proceedings of WSDM*, pages 13–22, 2013.


Jahna Otterbacher.
'helpfulness' in online communities: A measure of message quality.
In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 955–964, New York, NY, USA, 2009. ACM.