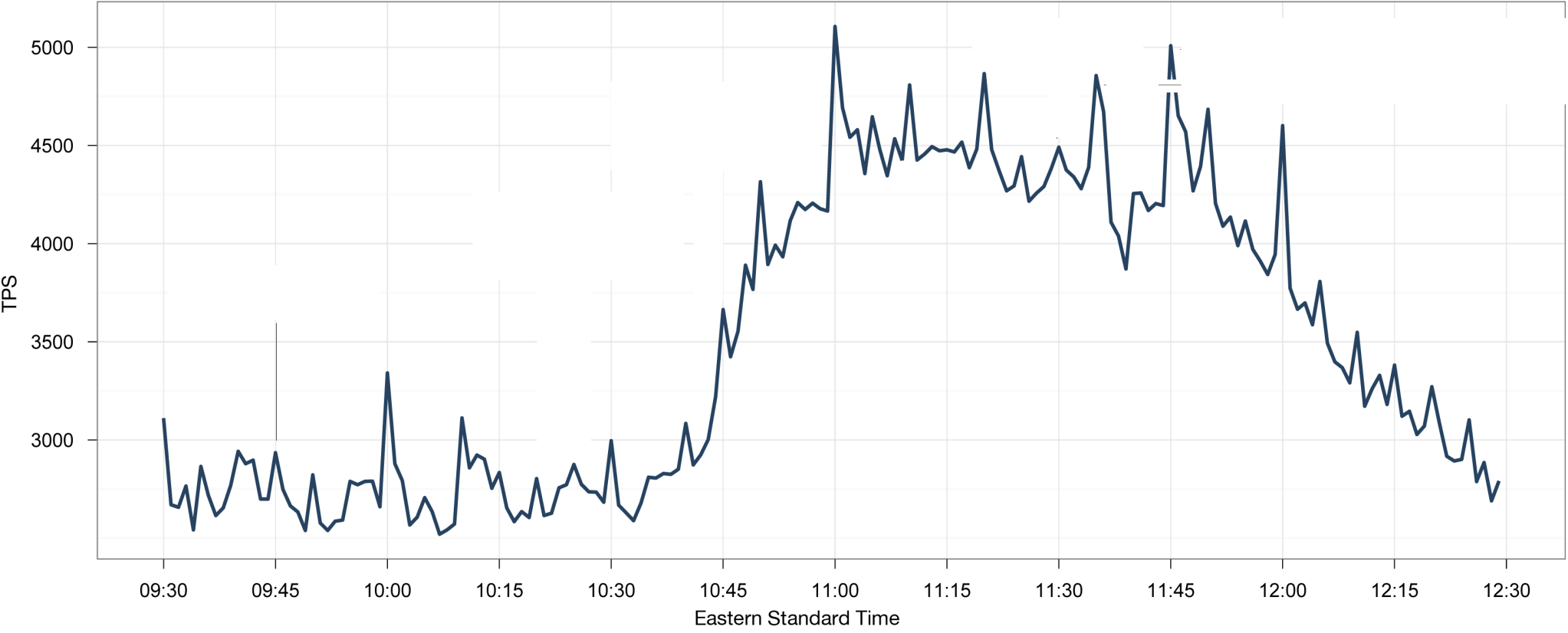


# TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration

Adam Marcus, Michael Bernstein, Osama Badar,  
David Karger, Sam Madden, Rob Miller

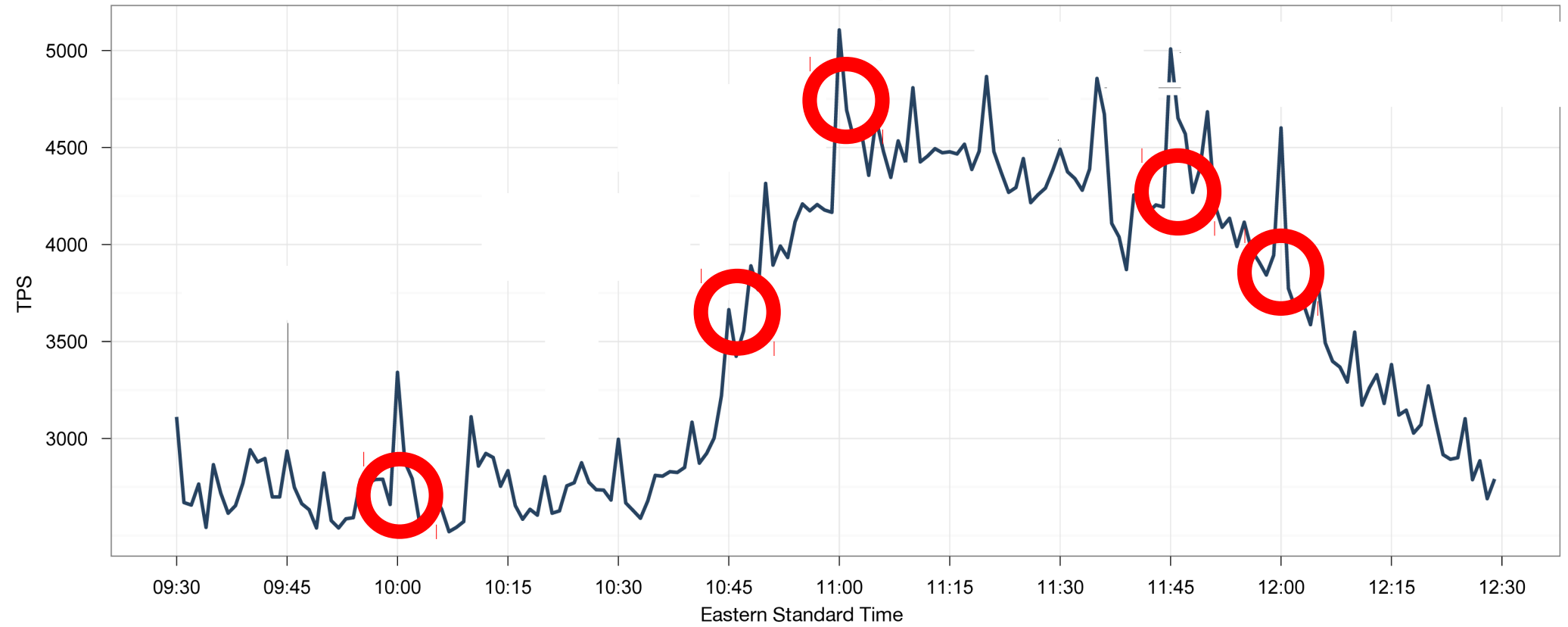
MIT CSAIL

# Tell stories using Twitter

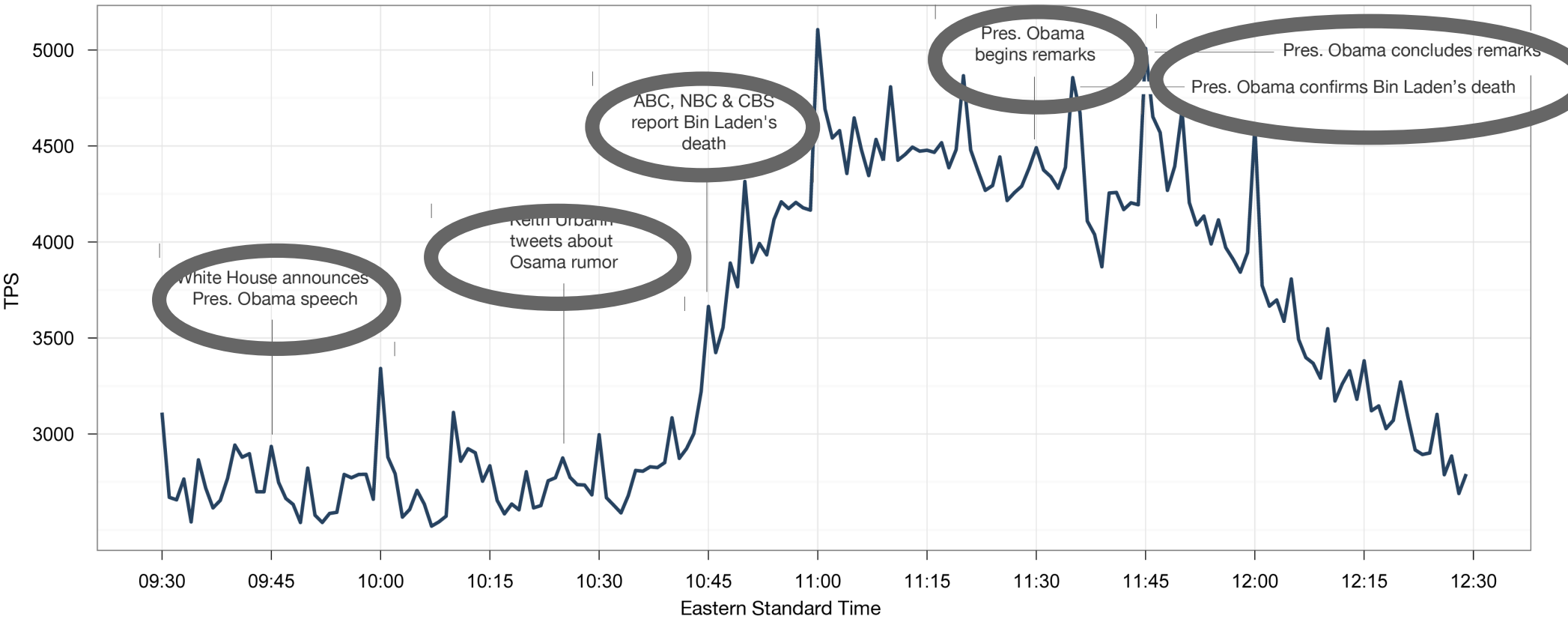




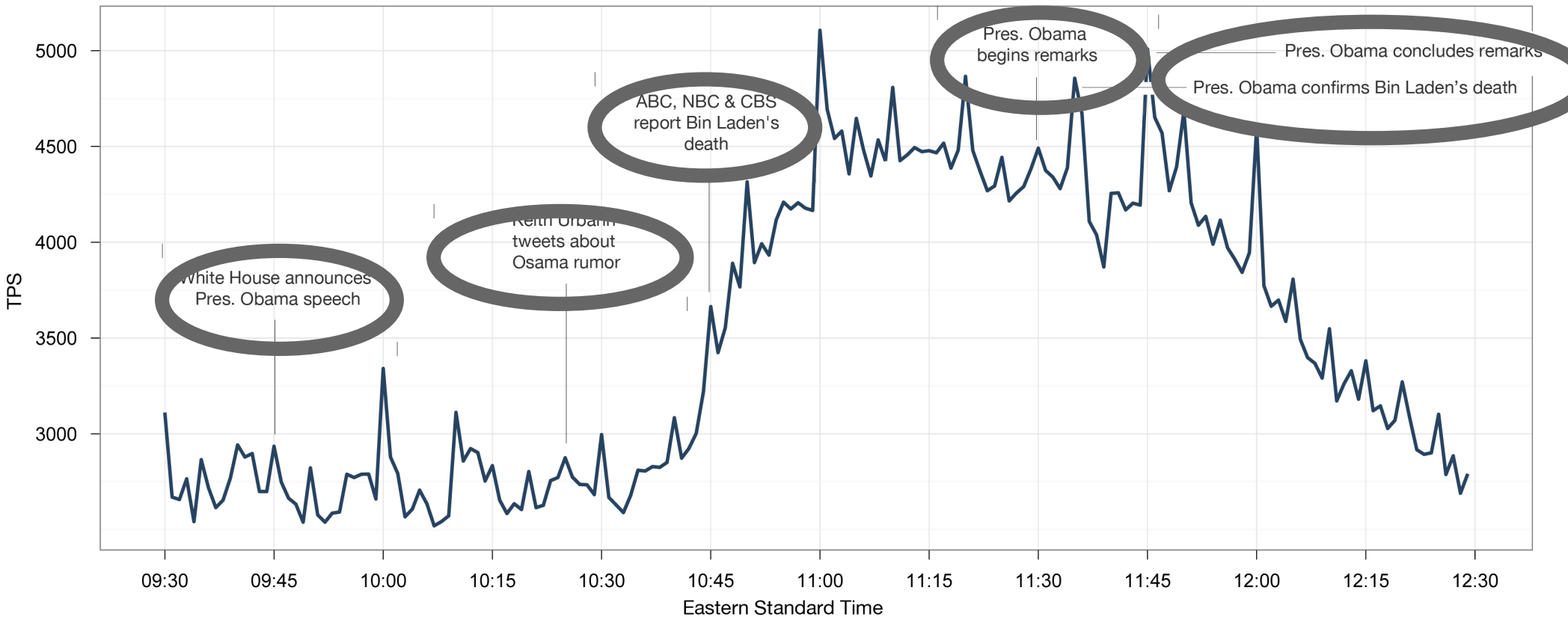




Identify events of interest



Identify events of interest  
Label events

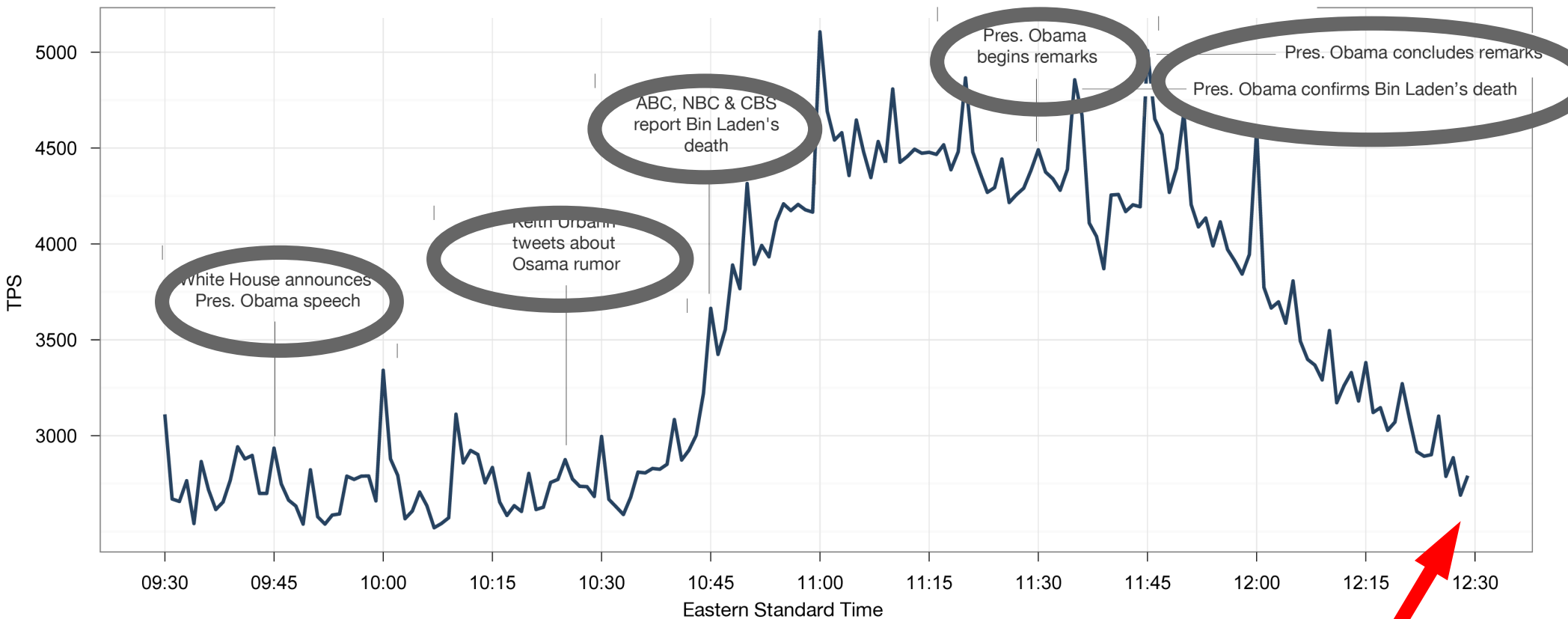


Identify events of interest

Label events

Add context: sentiment? location?





Identify events of interest

Label events

Add context: sentiment? location?

Streaming

## Tweets Per Second: Evening of May 1, 2011



**TwitInfo automates the process of telling stories using Twitter**

Identify events of interest

Label events

Add context: sentiment? location?

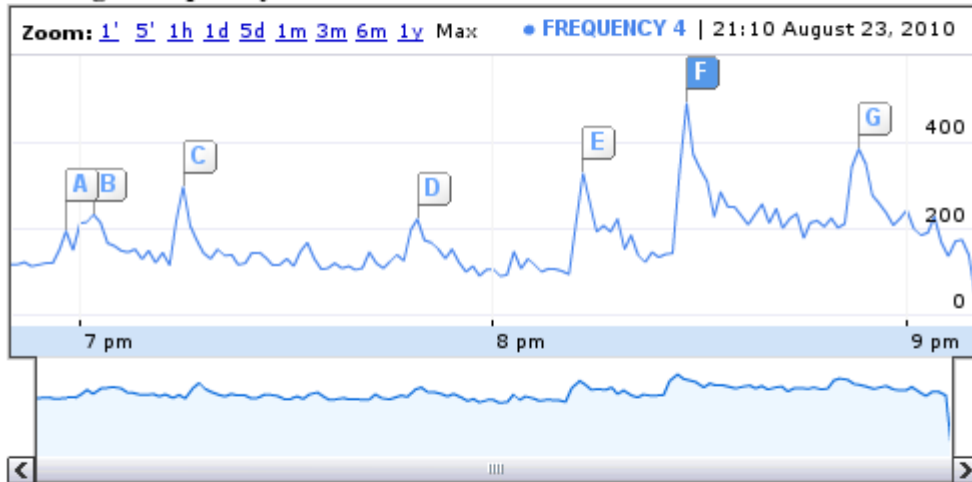
Streaming

## august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool

Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

### Message Frequency



### Tweet Map



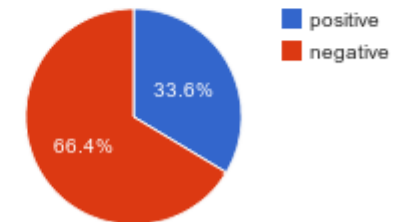
### Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;); <http://twurl.nl/knx4l>

### Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

### Overall Sentiment

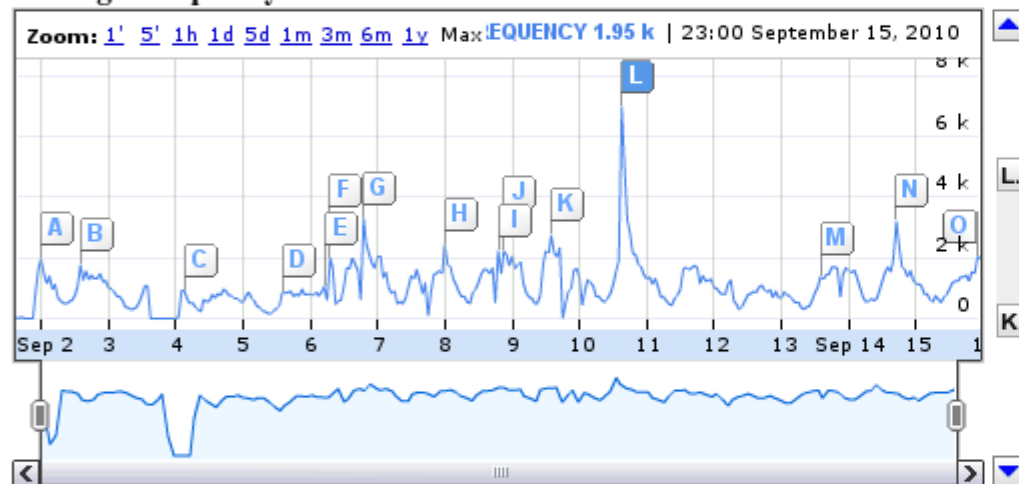


# barack obama

Keywords: obama

Event dates: Sept. 1, 2010, 3 p.m. - Sept. 16, 2010, midnight

## Message Frequency



## Tweet Map



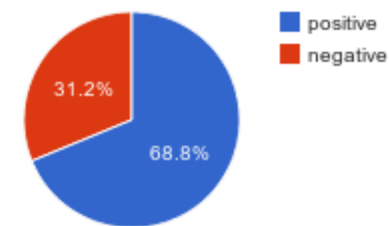
## Relevant Tweets

- 
**Goolsbee Said to Be Picked to Head White House Economic Panel: President Barack Obama will appoint Austan Goolsbee...**  
<http://bit.ly/birIPH>
- 
**Top Adviser to Lead Panel on Economy - NYTimes.com: President Obama on Friday will promote a longtime economic adv...**  
<http://bit.ly/bklmF1>
- 
**WASHINGTON — President Barack Obama has chosen one of**

## Popular Links

- <http://bit.ly/czJdxP> (cited by 22)
- <http://lat.ms/dcig1f> (cited by 4)
- <http://bit.ly/9XzAyd> (cited by 4)

## Overall Sentiment

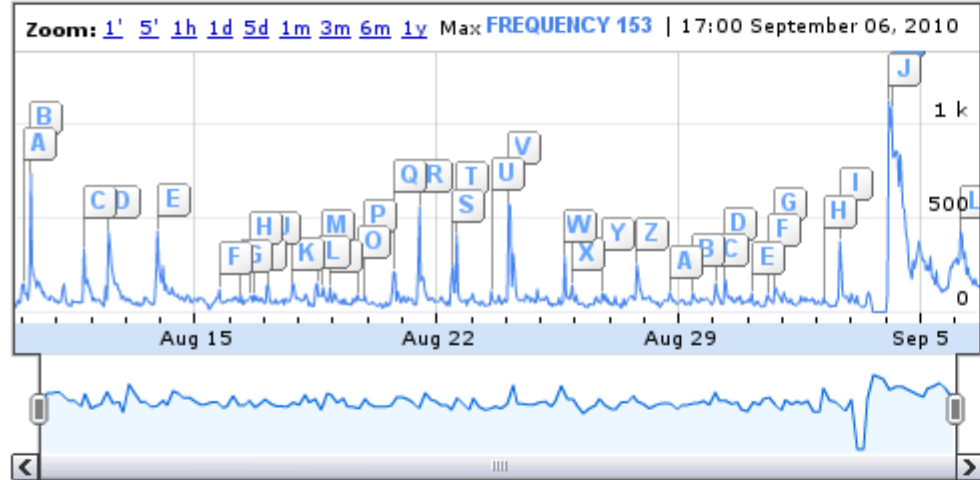


earthquakes

Keywords: earthquake

Event dates: Aug. 9, 2010, 7:18 p.m. - Sept. 6, 2010, 6 p.m.

Message Frequency



- Enter a filter term
- K. earthquake, christchurch, zealand, thoughts, prayers  
[Label this event](#)  
2010-9-4
  - J. earthquake, christchurch, 7.0, zealand, nz  
[Label this event](#)  
2010-9-4
  - I. earthquake, 3.3, felt, pablo, san  
[Label this event](#)  
2010-9-2

Relevant Tweets

RT @mariacelina: This morning, my thoughts & prayers go out to those affected by the earthquake that recently hit Christchurch, New Zealand.

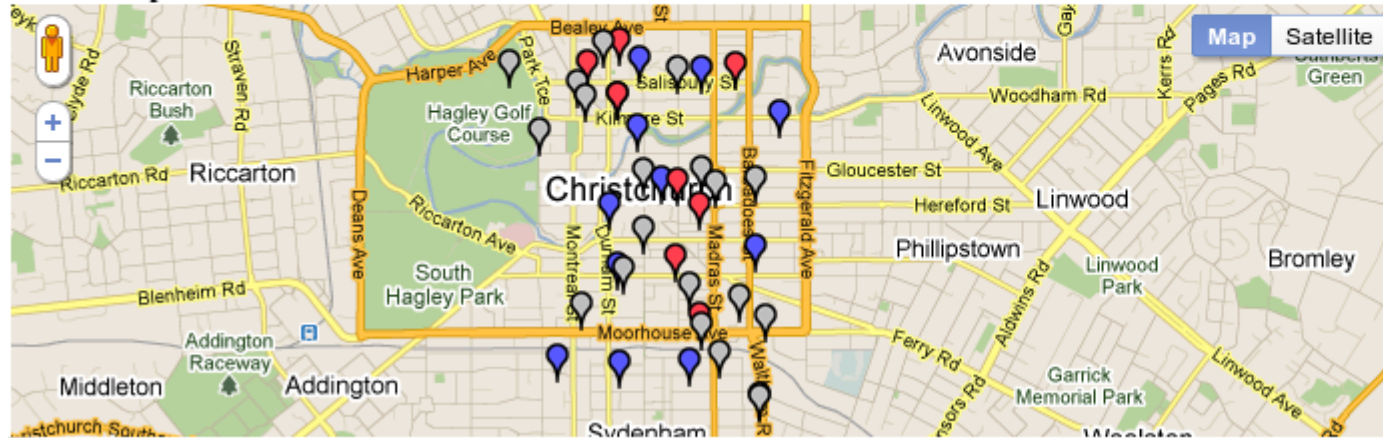
RT @KennyHamilton: Just heard about the earthquake in New Zealand...my thoughts and prayers go out to everyone in that country. I hope ...

RT @EvermoreBand: Thoughts and prayers with the people of Christchurch. Massive earthquake there, and no power or water for the whole city.

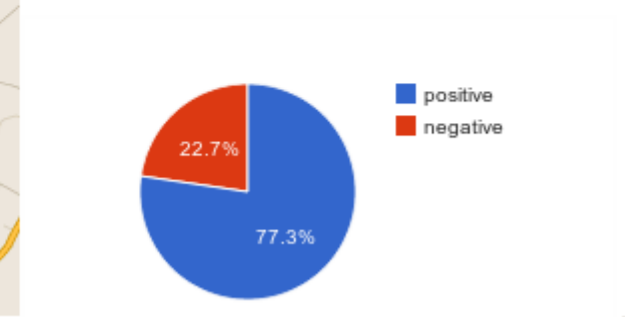
Popular Links

- <http://on.cnn.com/8YDPyG> (cited by 11)
- <http://bit.ly/ddWgtK> (cited by 6)
- <http://bit.ly/aFh7RM> (cited by 5)

Tweet Map



Overall Sentiment



# TwitInfo

event detection

event labeling

context

streaming algorithms

# TwitInfo

event detection

event labeling

context

streaming algorithms

sentiment correction algorithm

evaluation: news consumers

evaluation: pulitzer prize journalist

## State of the Union 2011

Tweet

(163,911 Tweets from 86,429 people)

### Search and Filter Tweets

(enter search keywords here)

Basic Filters:  
(click to toggle)

Advanced Filters:

### Video



### Tweets

### People

(from minute 7) Sort by: RTed Cou

Text	Date	RTed
<b>cspan</b> President Obama delivers his State of the Union address. <a href="http://twitpic.com/3tfolo">http://twitpic.com/3tfolo</a>	1/26/2011 02:19:05	RTed 7
<b>jezebeljenna</b> Friend in DC who works at a liquor store texted that Biden came by for a bottle of wine on his way to the SOTU. Party on, Joe.	1/26/2011 02:18:34	RTed 6
<b>seasonothebitch</b> You know why people had good jobs for life? The #labor movement. #classwar #sotu	1/26/2011 02:18:42	RTed 5
<b>jnicead1</b> Obama is the greatest orator since MLK and Kennedy...	1/26/2011 02:19:15	RTed 4
<b>whitehouse</b> "The rules have changed. In a single generation, revolutions in technology have transformed the way we live, work and do business." #sotu	1/26/2011 02:18:59	RTed 3
<b>mamichula1399</b> Yep, my vagina shriveled up and died. #sotu	1/26/2011 02:19:15	RTed 2

### Topic Sections

### Tweet Volume Over Time



### Sentiment Over Time

Positive Negative Controversial Neutral

### Key Words Over Time

congratulating largest biomedical winner kathy tsa simplification domestic earmarks troops forbidden

[Andre et al. UIST2007]      [Diakopoulos et al. VAST 2010]  
 [Diakopoulos+Shamma CHI2010]    [Dork et al. InfoVis2010]  
 [Go et al. 2010]      [Leskovec et al. KDD2009]  
 [Shamma et al. CSCW2010+2011]



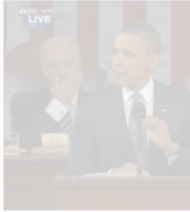
State of the Union 2011

Tweet

Search and Filter Twee

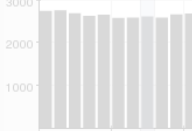
Basic Filters: (click to toggle) Quotes No Quote

Video



Topic Sections

Tweet Volume Over



Sentiment Over Time



Key Words Over Time

congratulating lar

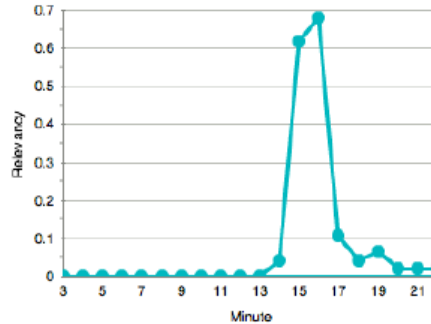
Statler

Debate

Inauguration

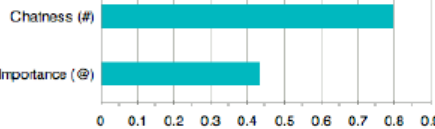


remaking's TF/IDF score over the inauguration  
Current Minute: 13



Click on a term to see its frequency graph  
(The top term's graph is displayed to the left.)

- remaking
- dust
- scripture
- greatness
- pick
- childish
- begin
- senior
- equal
- measure
- passed
- dogs
- failed
- score
- blogs
- criticize
- data
- stood
- capacity



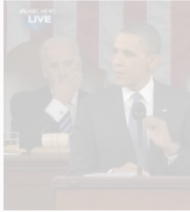
[Andre et al. UIST2007] [Diakopoulos et al. VAST 2010]  
 [Diakopoulos+Shamma CHI2010] [Dork et al. InfoVis2010]  
 [Go et al. 2010] [Leskovec et al. KDD2009]  
 [Shamma et al. CSCW2010+2011]

## State of the Union 2011

### Search and Filter Twee

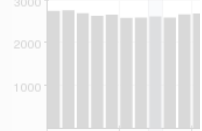
Basic Filters: (click to toggle)  
 Quotes No Quote

### Video



### Topic Sections

### Tweet Volume Over



### Sentiment Over Time



### Key Words Over Time

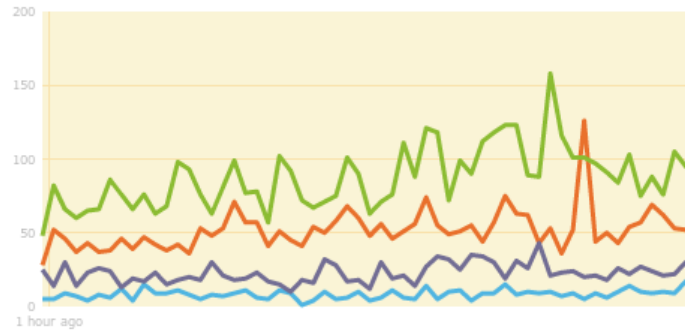
congratulating lar

## Statler

Debate Inauguration

### President Barack Obama 2009 Inauguration ...

★★★★★



remaking's TFIDF score over the inauguration  
 Current Minute: 13

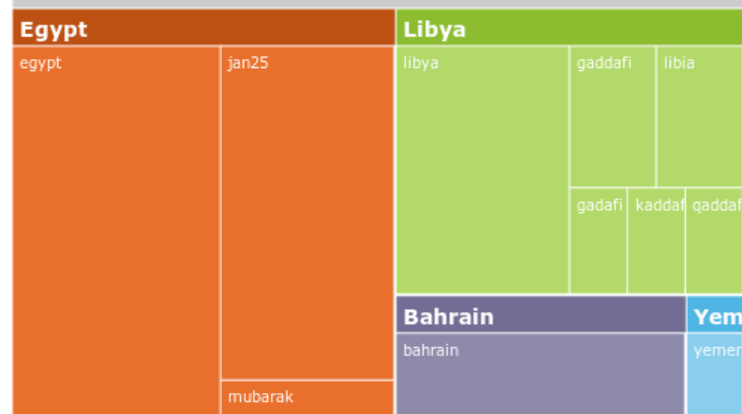
Click on a term to see its frequency graph  
 (The top term's graph is displayed to the left.)

TOTAL TWEETS TODAY AVERAGE (TWEETS/MINUTE)

Egypt	<b>33069</b>	<b>60</b>
Yemen	<b>5898</b>	<b>8</b>
Libya	<b>77626</b>	<b>78</b>
Bahrain	<b>12264</b>	<b>29</b>

- ebrahim\_khalil: RT @almameeen: <http://i.mp/epoJUo> تصويت : هل يسعى تلفزيون البحرين ربح الفقة الطقبة ؟ #bahrain #lulu #feb14 ريبويت
- yusufalkhan: RT @kamazaki050: #Bahrain له مزعجين من وصال يعني لو كان مالي احتلال للمستشفى ... و تحريض و ترهيب للصفار و الكبار كان الناس لشكوا و الا كل
- ZoozFadhul: RT @LiveBahrain: #Bahrain #Feb14 #lulu #UniteBH #LiveBahrain شارع الملك فيصل باتجاه دوار الزاوية معلق حاليا
- fredwillie460: @CNNLive US troops in #Bahrain might like to make the choice were they spend their \$'s <http://www.twitlonger.com/show/95t1ih>
- ebrahim\_khalil: RT @almameeen: <http://i.mp/epoJUo> تصويت : هل يسعى تلفزيون البحرين ربح الفقة الطقبة ؟ #bahrain #lulu #feb14 ريبويت
- yusufalkhan: RT @kamazaki050: #Bahrain له مزعجين من وصال يعني لو كان مالي احتلال للمستشفى ... و تحريض و ترهيب للصفار و الكبار كان الناس لشكوا و الا كل

### Keywords Distribution

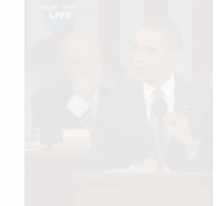


[Andre et al. UIST2007] [Diakopoulos et al. VAST 2010]  
 [Diakopoulos+Shamma CHI2010] [Dork et al. InfoVis2010]  
 [Go et al. 2010] [Leskovec et al. KDD2009]  
 [Shamma et al. CSCW2010+2011]

Search and Filter Twee

Basic Filters: Quotes  
(click to toggle) No Quote

Video



Topic Sections



Key Words Over Time

Statler  
 Debate Inauguration

**President Barack Obama 2009 Inauguration ...**  
 ★ ★ ★ ★ ★  
 remaking's TFIDF score over the inauguration  
 Current Minute: 13

remaking's TFIDF score over the inauguration  
 Current Minute: 13

Click on a term to see its frequency graph  
 (The top term's graph is displayed to the left.)

	TOTAL TWEETS TODAY	AVERAGE (TWEETS/MINUTE)
Egypt	33069	60
Yemen	5898	8
Libya	77626	78
Bahrain	12264	29

**TwitInfo is a streaming storytelling layer that complements these visualizations**

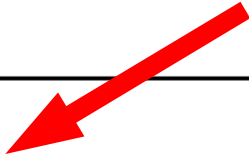
shrahim\_khalil: RT @almameeen: <http://i.mp/epo116> ...  
 fredwillie460: @CNNLive US troops in #Bahrain might like to make the choice were they spend their \$'s <http://www.twitlonger.com/show/95t11h>  
 shrahim\_khalil: RT @almameeen: <http://i.mp/epo116> ...  
 youfalkhan: RT @kamazaki050: #Bahrain ...



[Andre et al. UIST2007] [Diakopoulos et al. VAST 2010]  
 [Diakopoulos+Shamma CHI2010] [Dork et al. InfoVis2010]  
 [Go et al. 2010] [Leskovec et al. KDD2009]  
 [Shamma et al. CSCW2010+2011]

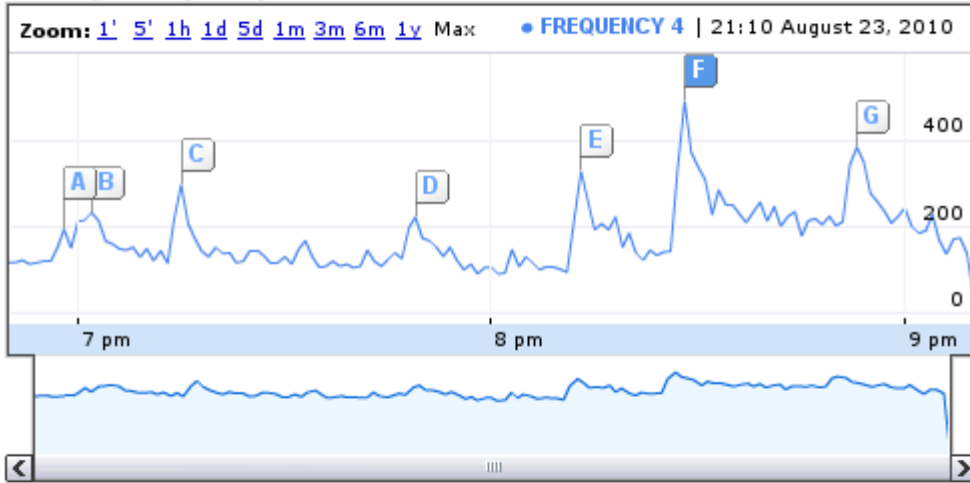
**ui walkthrough**  
algorithm design  
study/evaluation  
discussion

august 23 manchester city vs. liverpool



Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool  
 Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

Message Frequency



- Enter a filter term
- G. liverpool, city, 3-0, kalah, tevez  
[Label this event](#)  
 2010-8-23
  - F. liverpool, city, tevez, manciny, 3-0  
[Label this event](#)  
 2010-8-23
  - E. liverpool, city, manciny, 2-0, man  
[Label this event](#)  
 2010-8-23

Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knx4l>

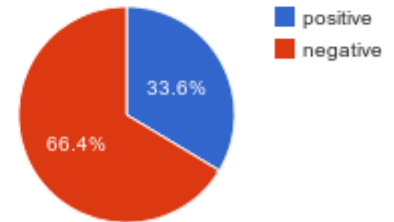
Popular Links

- <http://bit.ly/aLlUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

Tweet Map



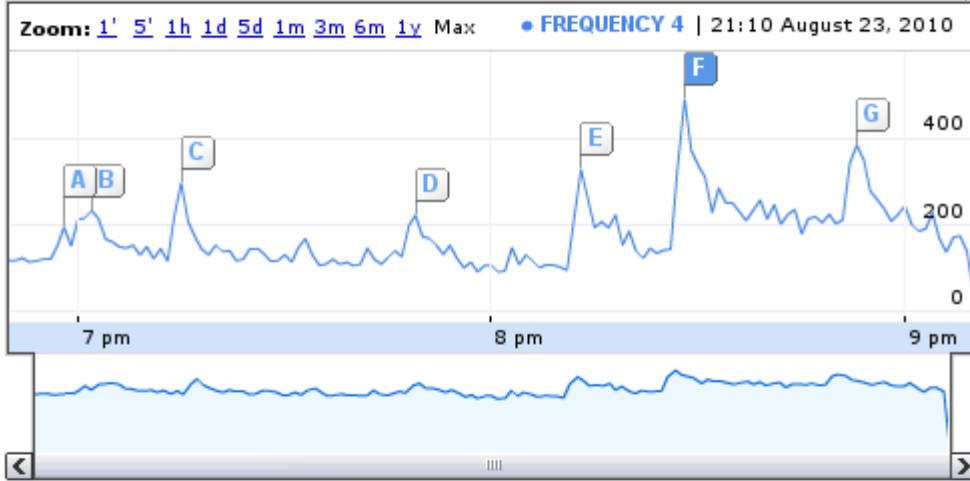
Overall Sentiment



# august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, mancity, liverpool  
 Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

## Message Frequency



## Tweet Map



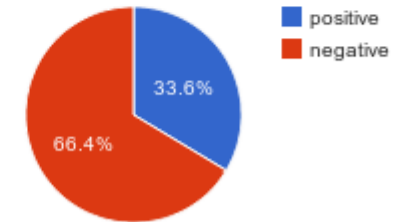
## Relevant Tweets

- Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
- Man City 3-0 Liverpool (Tevez pen)
- penalty for City. And Tevez makes it 3-0 for ManCity
- Mancity vs livpool 3-0 tevez gooolllllll
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knx4l>

## Popular Links

- <http://bit.ly/aLlUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

## Overall Sentiment

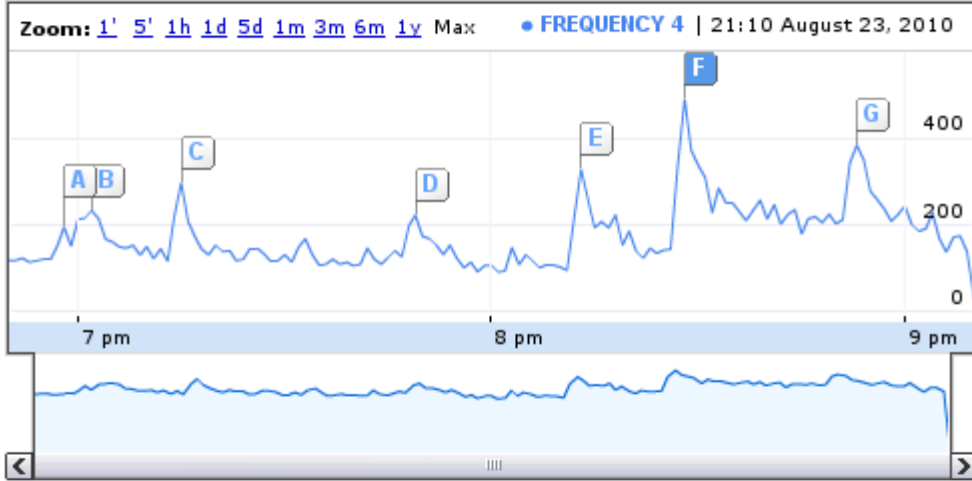




# august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool  
Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

## Message Frequency



## Tweet Map



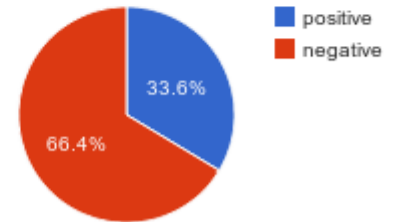
## Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knx4l>

## Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

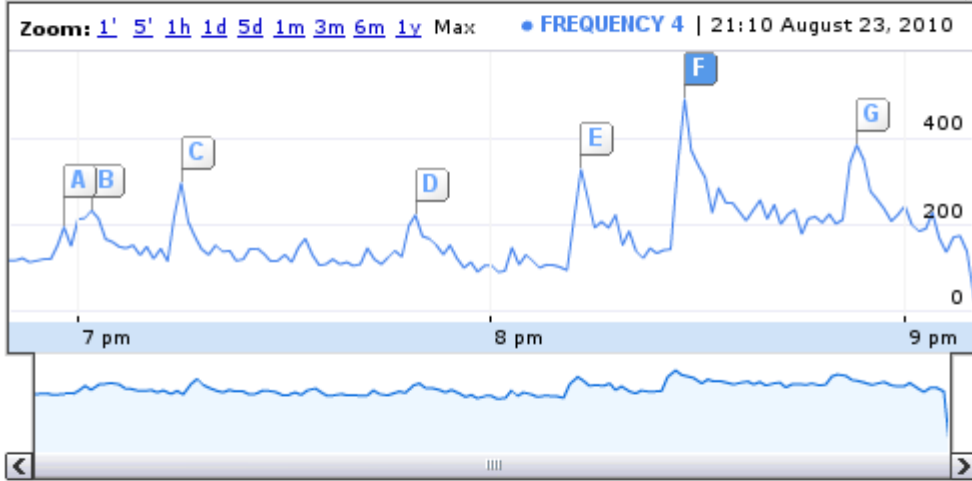
## Overall Sentiment



# august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool  
 Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

## Message Frequency



## Tweet Map



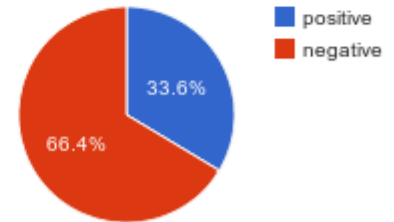
## Relevant Tweets

- 
- 
- 
- 
- 

## Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

## Overall Sentiment

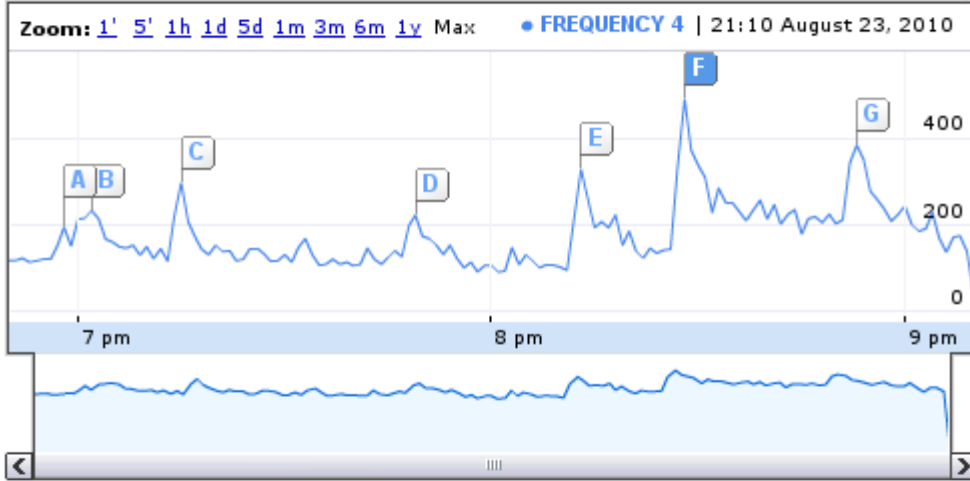




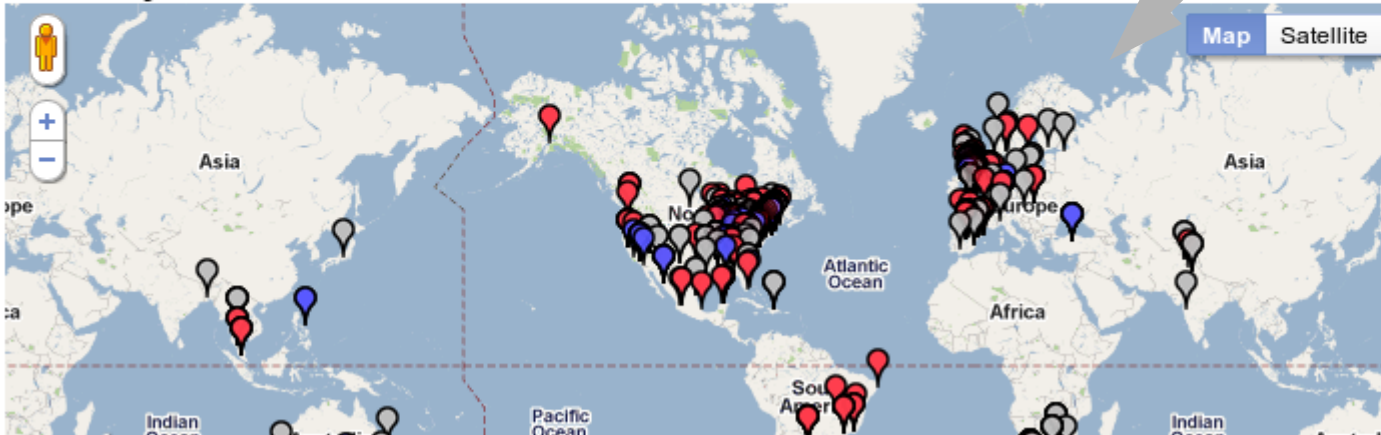
## august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool  
 Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

### Message Frequency



### Tweet Map



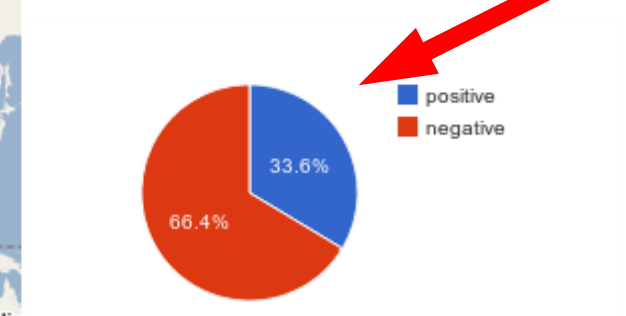
### Relevant Tweets

- Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
- Man City 3-0 Liverpool (Tevez pen)
- penalty for City. And Tevez makes it 3-0 for ManCity
- Mancity vs livpool 3-0 tevez gooolllllll
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knx4l>

### Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

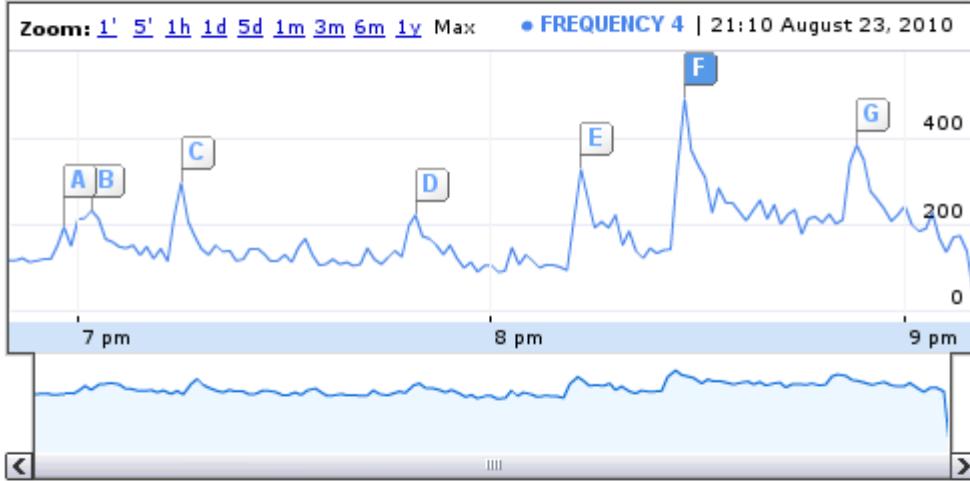
### Overall Sentiment



## august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, manciny, liverpool  
 Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

### Message Frequency



### Tweet Map



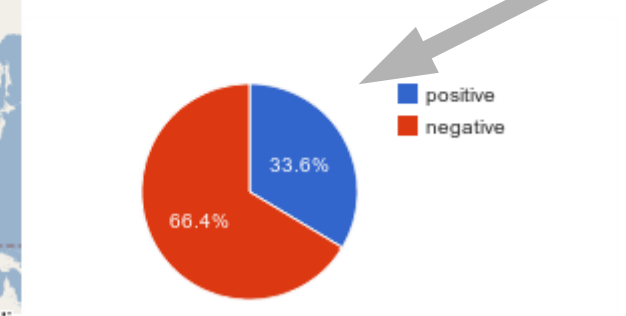
### Relevant Tweets

- Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
- Man City 3-0 Liverpool (Tevez pen)
- penalty for City. And Tevez makes it 3-0 for ManCity
- Mancity vs livpool 3-0 tevez gooolllllll
- BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knky...>

### Popular Links

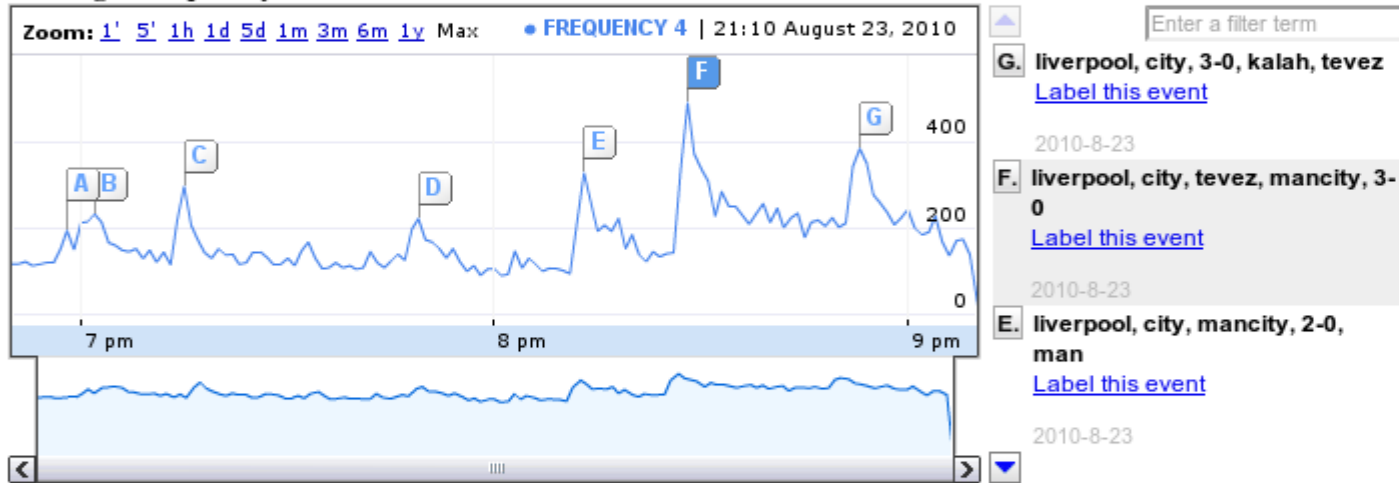
- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

### Overall Sentiment

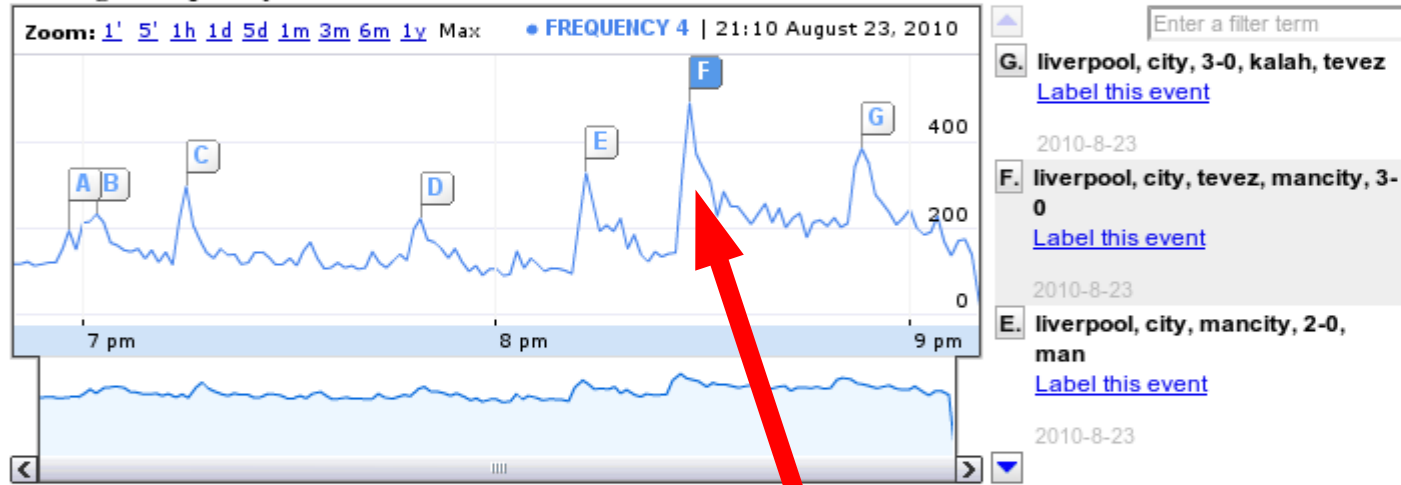


ui walkthrough  
**algorithm design**  
study/evaluation  
discussion

## Message Frequency

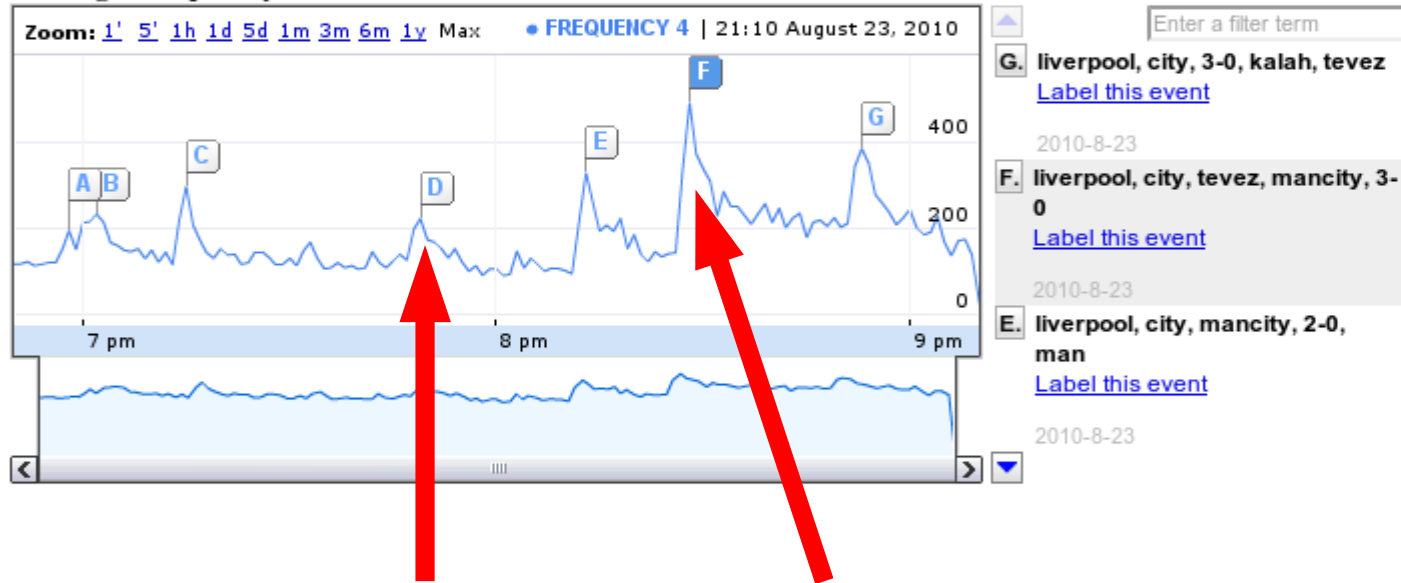


## Message Frequency



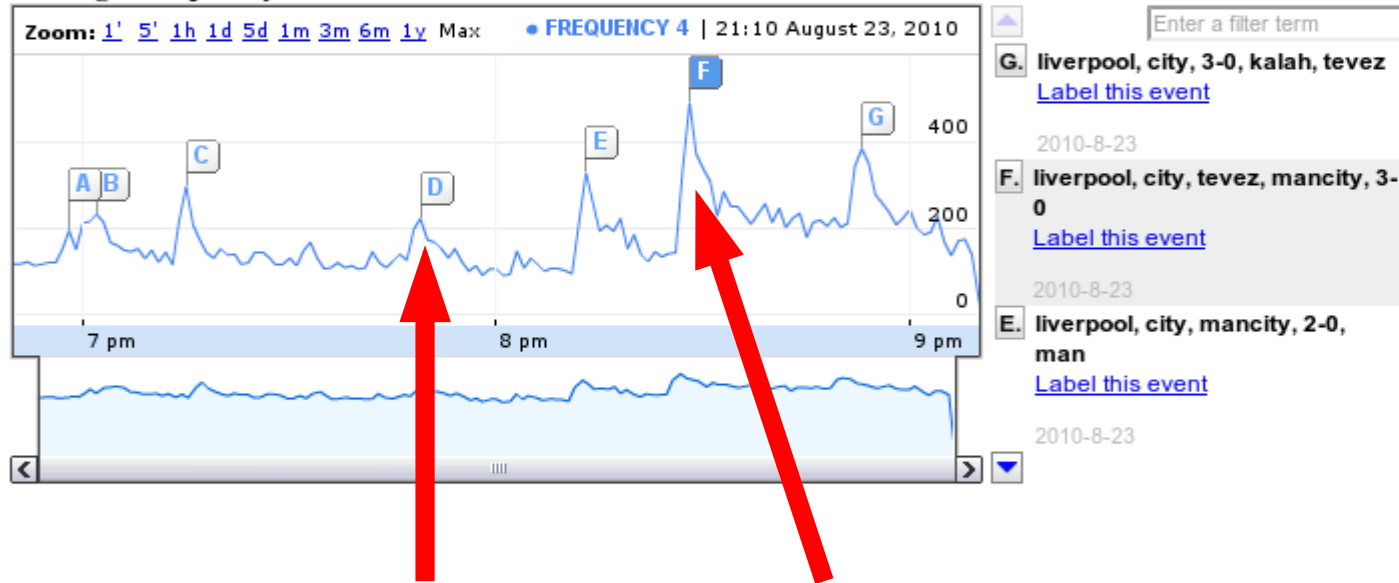
Large

## Message Frequency



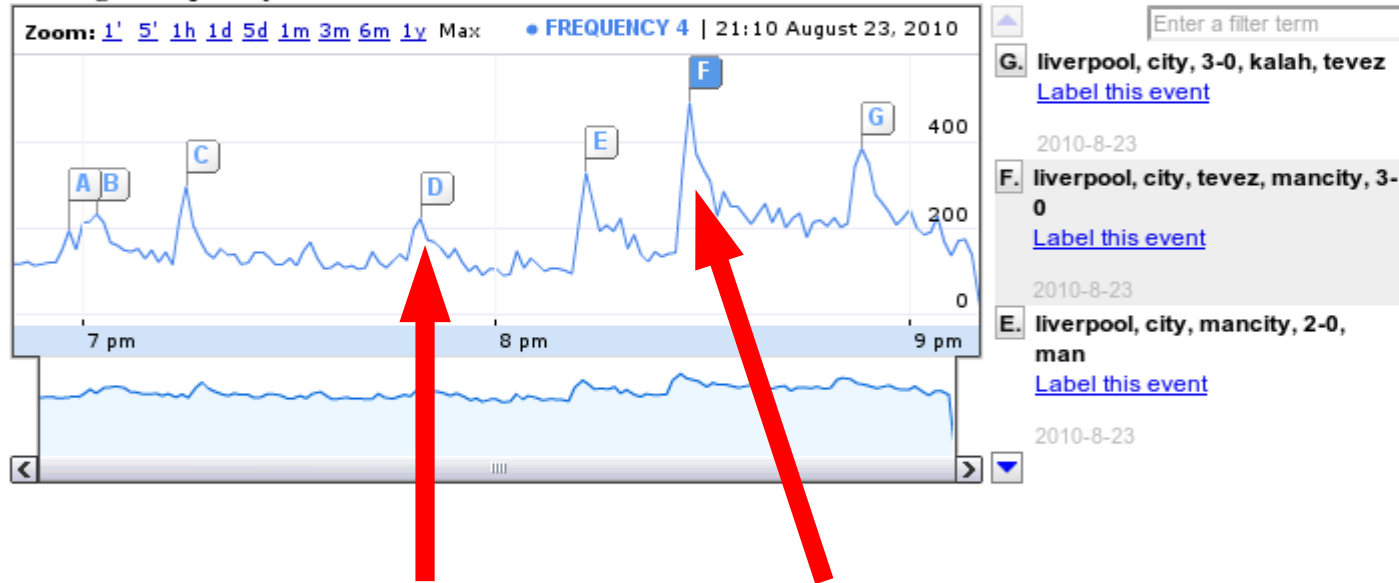
Relatively Large

## Message Frequency



# Streaming

## Message Frequency

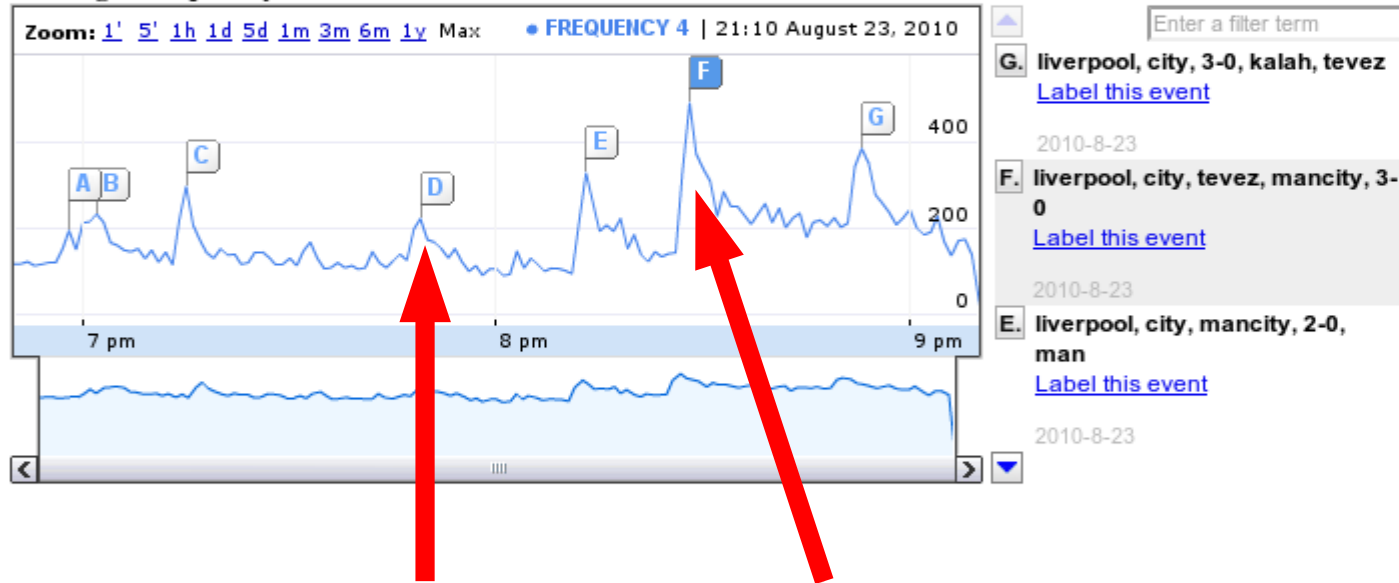


Relatively Large

Streaming  
mean deviation



## Message Frequency

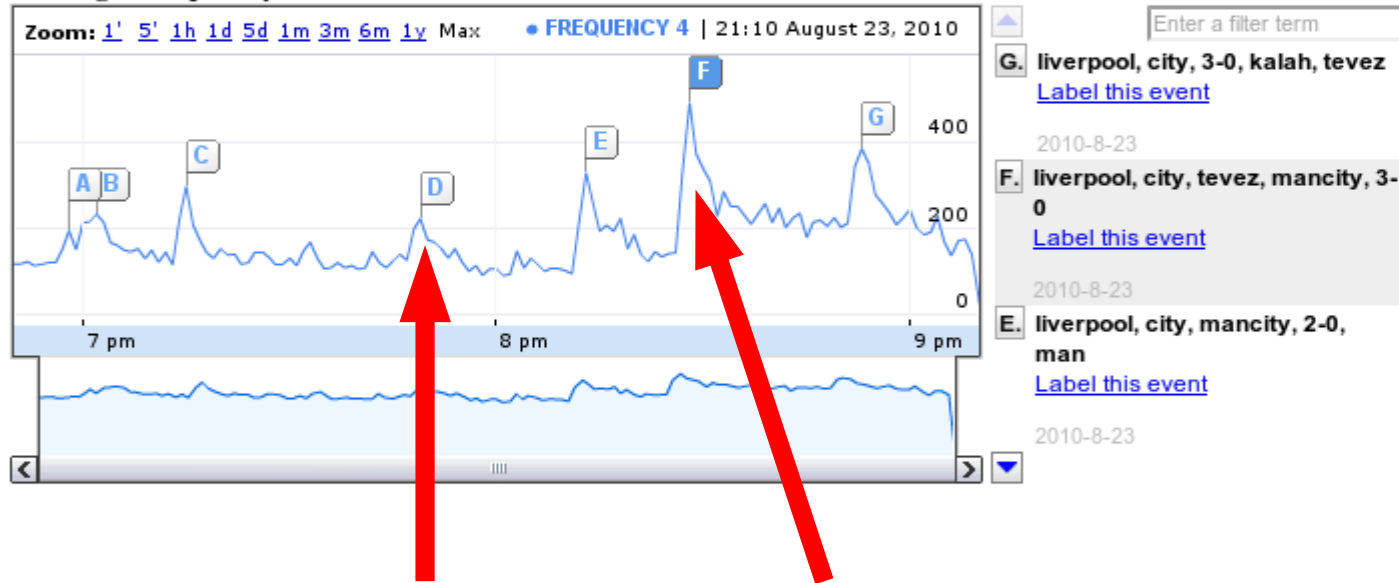


Relatively Large

Adaptive

Streaming  
mean deviation

## Message Frequency

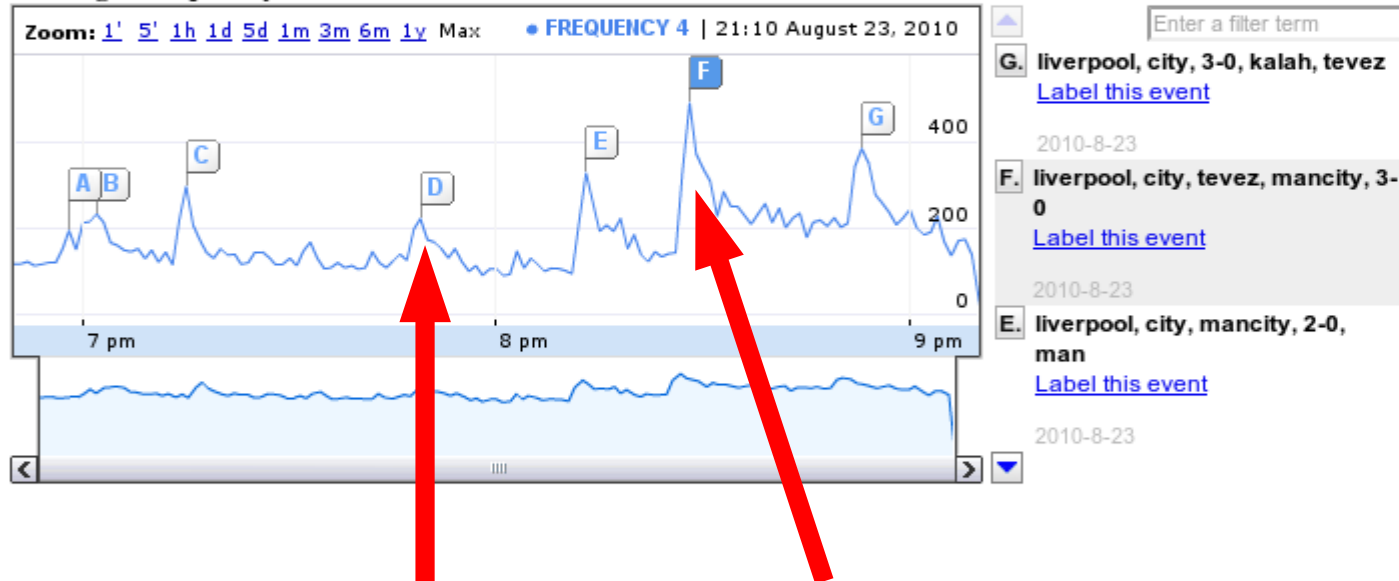


Adaptive

Streaming

Exponentially weighted moving mean deviation

## Message Frequency



Relatively Large

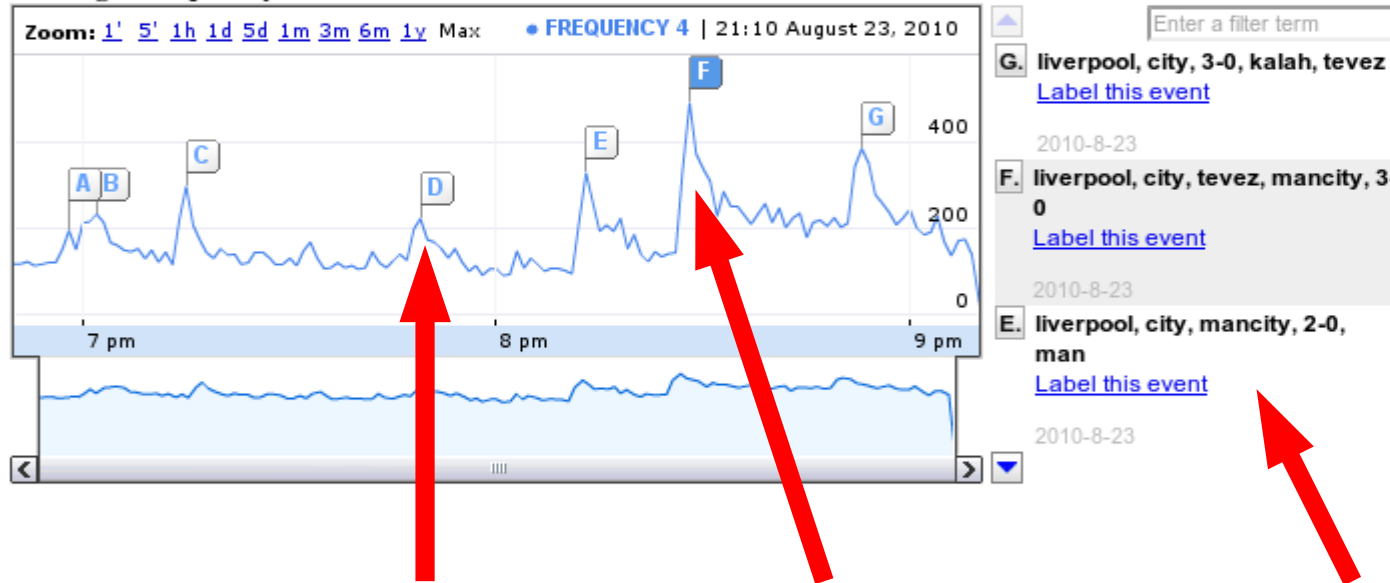
Adaptive

Streaming

Exponentially weighted moving mean deviation

RFC 2988: Computing  
TCP's retransmission timer

## Message Frequency



Relatively Large

TF-IDF

Adaptive

Streaming

Exponentially weighted moving mean deviation

RFC 2988: Computing  
TCP's retransmission timer

## Message Frequency



Precision: 80-100%

Recall: 80-100%

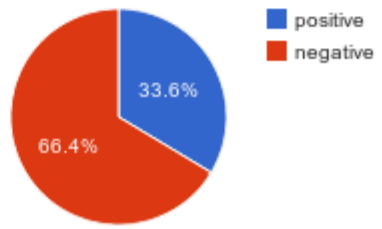
Adaptive

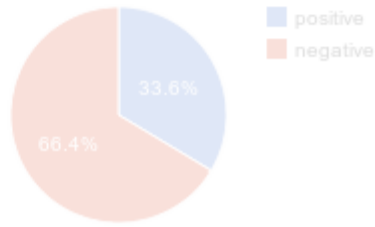
Streaming

Exponentially weighted moving mean deviation

RFC 2988: Computing

TCP's retransmission timer

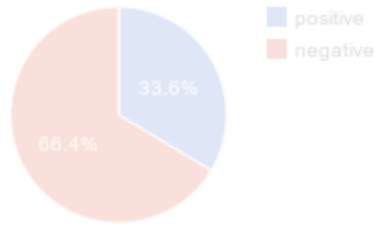




## Negative Tweets (500 tweets)



Classified as Negative  
(50 tweets)

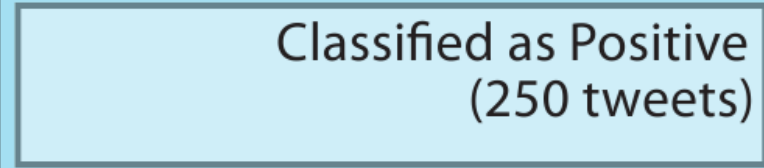


**Negative Tweets**  
(500 tweets)



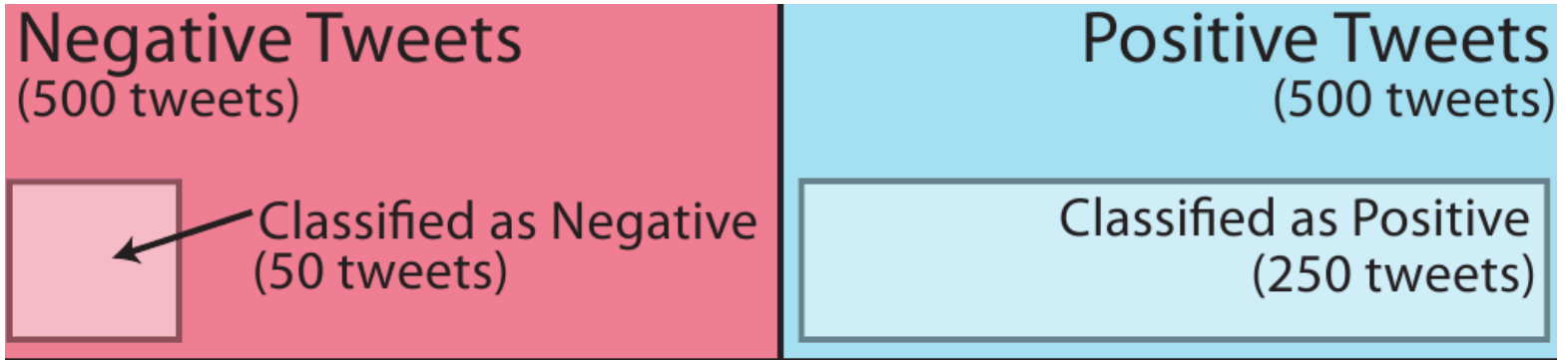
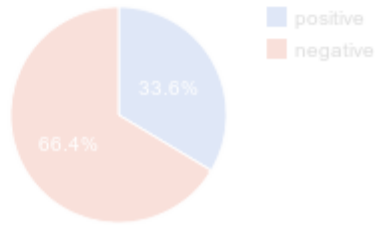
Classified as Negative  
(50 tweets)

**Positive Tweets**  
(500 tweets)

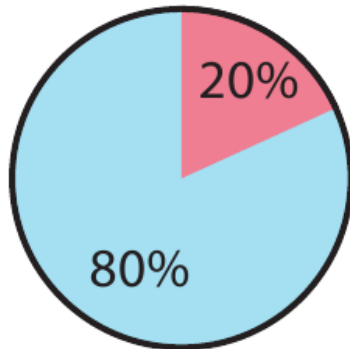


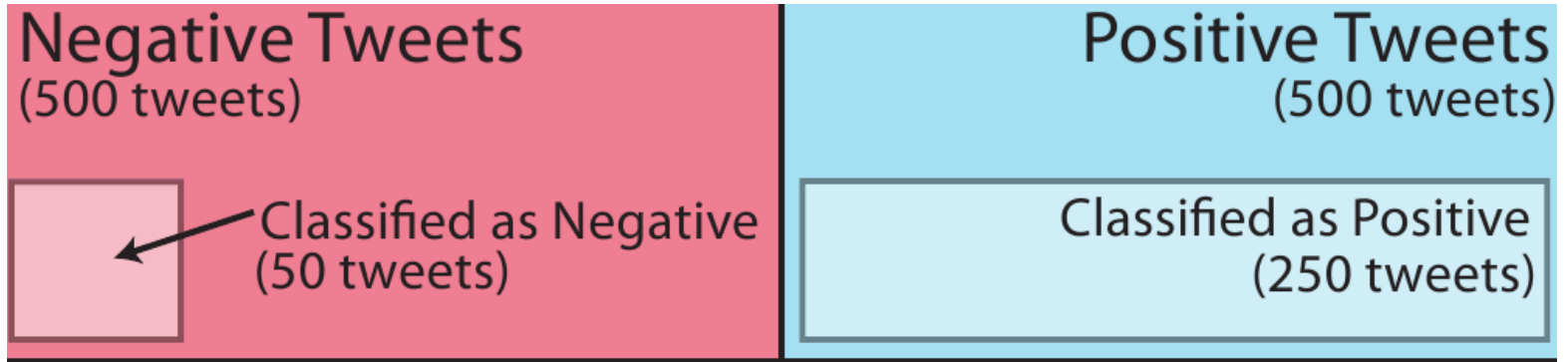
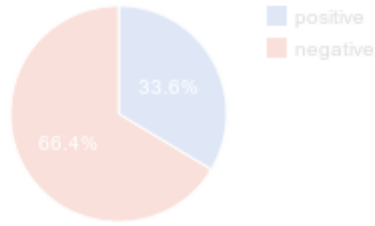
Classified as Positive  
(250 tweets)



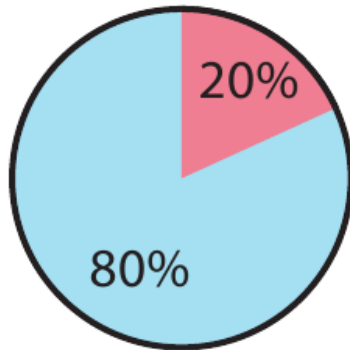


Uncorrected Output:

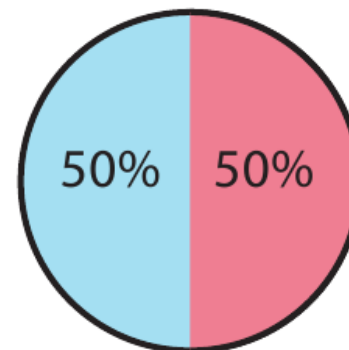




Uncorrected Output:



Recall-Normalized Output:



**MAYBE CUT!**

ui walkthrough  
algorithm design  
**study/evaluation**  
discussion

12 participants: four female/8 male  
first half: directed tasks  
second half: 5 min write article (soccer, Obama)

how do they use the interface?  
what can they learn from TwitInfo?

## Participant 6:

“After the peace talks, Obama traveled to the ASEAN conference and to NATO to work on issues in those parts of the world. He then spent the week dedicated to domestic economic issues. First he proposed a research tax break, then a \$50 billion investment in infrastructure, then the issue came up about whether he should keep some tax breaks that Bush had implemented, and he’s asking for some tax breaks from business, and these are generating some controversy because [. . . ]”

## Message Frequency



- Enter a filter term
- G. liverpool, city, 3-0, kalah, tevez**  
[Label this event](#)  
2010-8-23
  - F. liverpool, city, tevez, manccity, 3-0**  
[Label this event](#)  
2010-8-23
  - E. liverpool, city, manccity, 2-0, man**  
[Label this event](#)  
2010-8-23

## Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knkx4l>

## Popular Links

- <http://bit.ly/alLvUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

## Message Frequency



- Enter a filter term
- G.** liverpool, city, 3-0, kalah, tevez  
[Label this event](#)  
2010-8-23
  - F.** liverpool, city, tevez, manccity, 3-0  
[Label this event](#)  
2010-8-23
  - E.** liverpool, city, manccity, 2-0, man  
[Label this event](#)  
2010-8-23

## Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knkx4l>

## Popular Links

- <http://bit.ly/alLvUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

## Message Frequency



- G [liverp](#)  
[Label](#)
- 2010-
- F [liverp](#)  
[0](#)  
[Label](#)
- 2010-
- E [liverpool, city, manc](#)  
[city, 2-0,](#)  
[man](#)  
[Label this event](#)
- 2010-8-23

## Relevant Tweets

 Tevez... Nets it man! ManCity 3-0  
Liverpool... #mogbe

 Man Citv 3-0 Liverpool (Tevez

---

/. And Tevez makes  
ity

---

ool 3-0 tevez

---

COM - Update:  
y - Liverpool 3-0  
(Tevez 66&#039;) <http://twurl.nl/knkx4l>

# Split

## Popular Links

- <http://bit.ly/alLvUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)



## Message Frequency



Everyone

- G. liverp  
[Label](#)
- 2010-
- F. liverp  
0  
[Label](#)
- 2010-
- E. liverpool, city, mancify, 2-0,  
man  
[Label this event](#)
- 2010-8-23

## Relevant Tweets



Tevez... Nets it man! ManCity 3-0  
Liverpool... #mogbe



Man Citv 3-0 Liverpool (Tevez

---

/. And Tevez makes  
ity

---

ool 3-0 tevez

---

COM - Update:  
y - Liverpool 3-0  
(Tevez 66&#039;) <http://twurl.nl/knkx4l>

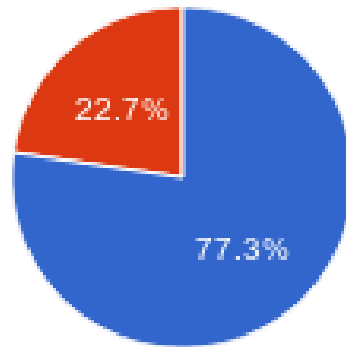
Split

## Popular Links

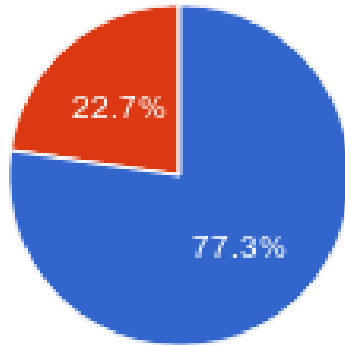
- <http://bit.ly/alLvUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)

News  
Junkies

# earthquakes



# earthquakes



“my thoughts and prayers go out to everyone in that country. I hope...”

# Journalist Interview

# Journalist Interview

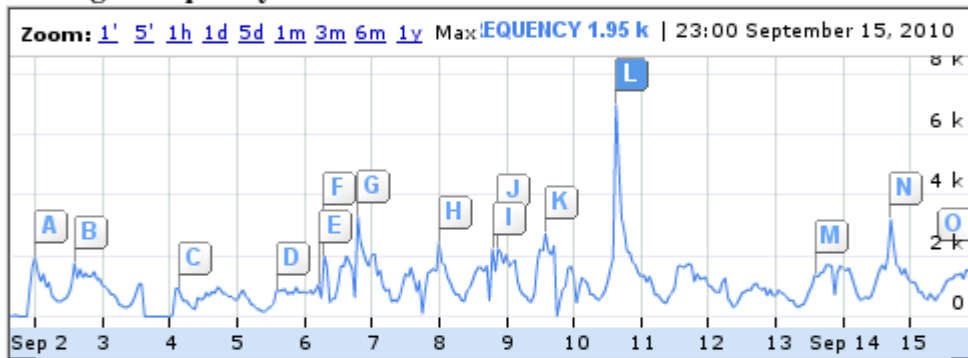
## backgrounding

barack obama

Keywords: obama

Event dates: Sept. 1, 2010, 3 p.m. - Sept. 16, 2010, midnight

### Message Frequency



# Journalist Interview

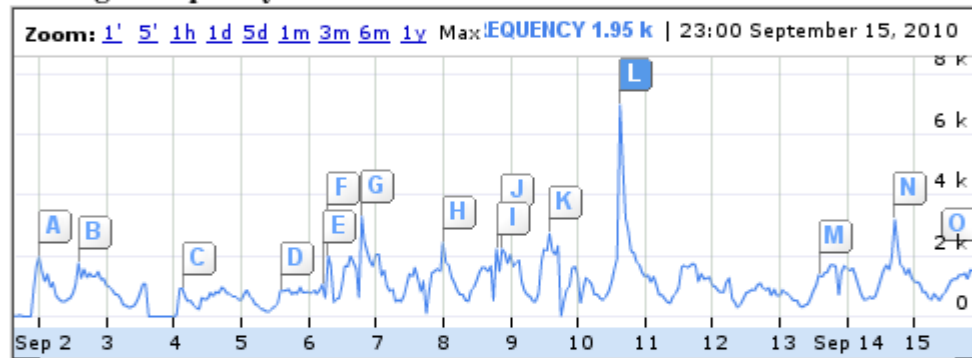
## backgrounding

barack obama

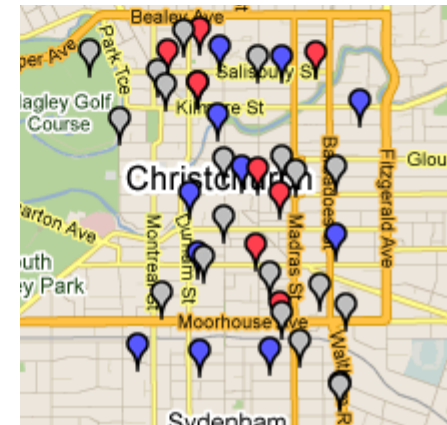
Keywords: obama

Event dates: Sept. 1, 2010, 3 p.m. - Sept. 16, 2010, midnight

### Message Frequency



## eyewitnesses



ui walkthrough  
algorithm design  
study/evaluation  
**discussion**

appropriate algorithms for the medium



appropriate algorithms for the medium

**other uses for streaming algorithms/signal processing**

appropriate algorithms for the medium

other uses for streaming algorithms/signal processing

**sentiment: caution!**

appropriate algorithms for the medium

other uses for streaming algorithms/signal processing

sentiment: caution!

**mixed-initiative storytelling**

# TwitInfo

<http://twitinfo.csail.mit.edu>

[marcua@csail.mit.edu](mailto:marcua@csail.mit.edu) / [@marcua](#)

Adam Marcus, Michael Bernstein, Osama Badar,  
David Karger, Sam Madden, Rob Miller



# Twitter data for fun and stories

# Twitter data for fun and stories

How can we tell stories with the tweet stream?

# Twitter data for fun and stories

How can we tell stories with the tweet stream?

How can we extract data from the tweet stream?



# Twitter data for fun and stories

How can we tell stories with the tweet stream?

TwitInfo

How can we extract data from the tweet stream?

# Twitter data for fun and stories

How can we tell stories with the tweet stream?

TwitInfo

How can we extract data from the tweet stream?

TweeQL

# TwitInfo

How can we tell stories with the tweet stream?

# TwitInfo

How can we tell stories with the tweet stream?

twitInfo

august 23 manchester city vs. liverpool **1**

Keywords: football, soccer, nfl, premier, league, premierleague, manchester city, mancity, liverpool.  
Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 25, 2010, 9:10 p.m.

Message Frequency



Tweet Map **3**



Relevant Tweets **4**

A list of relevant tweets is displayed. The first tweet is from a user with a purple profile picture, stating "I'm getting ready for the liverpool game. I'm so excited". The second tweet is from @football.Love, mentioning Sheikh Mansour bin Zayed and the ManCity vs Liverpool match. The third tweet is from Javier Mascherano, stating he refuses to face Manchester City. The fourth tweet is about online football gambling sites. The fifth tweet is a retweet from @wisnuwardhana about watching the ManCity vs Liverpool game.

Popular Links **5**

<http://bit.ly/PROVa> (cited by 4)  
<http://tinyurl.com/2d4s46d> (cited by 4)

Overall Sentiment **6**



# TwitInfo Uses

Automatically identify events  
e.g., goals, earthquakes

# TwitInfo Uses

Automatically identify events  
e.g., goals, earthquakes

Backgrounding  
e.g., Obama's last two weeks

# TwitInfo Uses

Automatically identify events  
e.g., goals, earthquakes

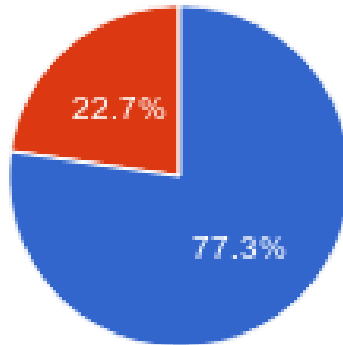
Backgrounding  
e.g., Obama's last two weeks

Identifying sources on the ground  
e.g., interviewing earthquake survivors

# Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”





# Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”

Location is evasive

# Cautionary Tweet Tales

Sentiment is tricky

e.g., “My warmest prayers go out to the people of Christchurch.”

Location is evasive

Spam is everywhere

# TwitInfo

Automated event detection makes tweet-based  
story-telling possible

# TwitInfo

Automated event detection makes tweet-based story-telling possible

Interfaces tell a story, people add context

# TwitInfo

Automated event detection makes tweet-based story-telling possible

Interfaces tell a story, people add context



# TweeQL

How can we extract data from the tweet stream?

# TweeQL

How can we extract data from the tweet stream?

“It's a balmy 89°C in Phoenix”

location=Phoenix, temperatureC=89

# TweeQL

How can we extract data from the tweet stream?

“It's a balmy 89°C in Phoenix”

location=Phoenix, temperatureC=89

“I'm starting to dig Obamacare!”

topic=Obama, sentiment=positive



# Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

# Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

Ad-hoc data processing

# Twitter data hacking is hard

- Learn the API
- Transform data
- Stuff it into a database

Ad-hoc data processing



TweeQL extracts data from tweets as they pass through the stream

TweeQL extracts data from tweets as they pass through the stream

Data extraction

e.g., location, sentiment, temperature, opencalais

# TweeQL extracts data from tweets as they pass through the stream

Data extraction

e.g., location, sentiment, temperature, opencalais

SQL-like queries

e.g., `SELECT location, text FROM twitter`

# TweeQL extracts data from tweets as they pass through the stream


Data extraction

e.g., location, sentiment, temperature, opencalais

SQL-like queries

e.g., `SELECT location, text FROM twitter`

stream, not table



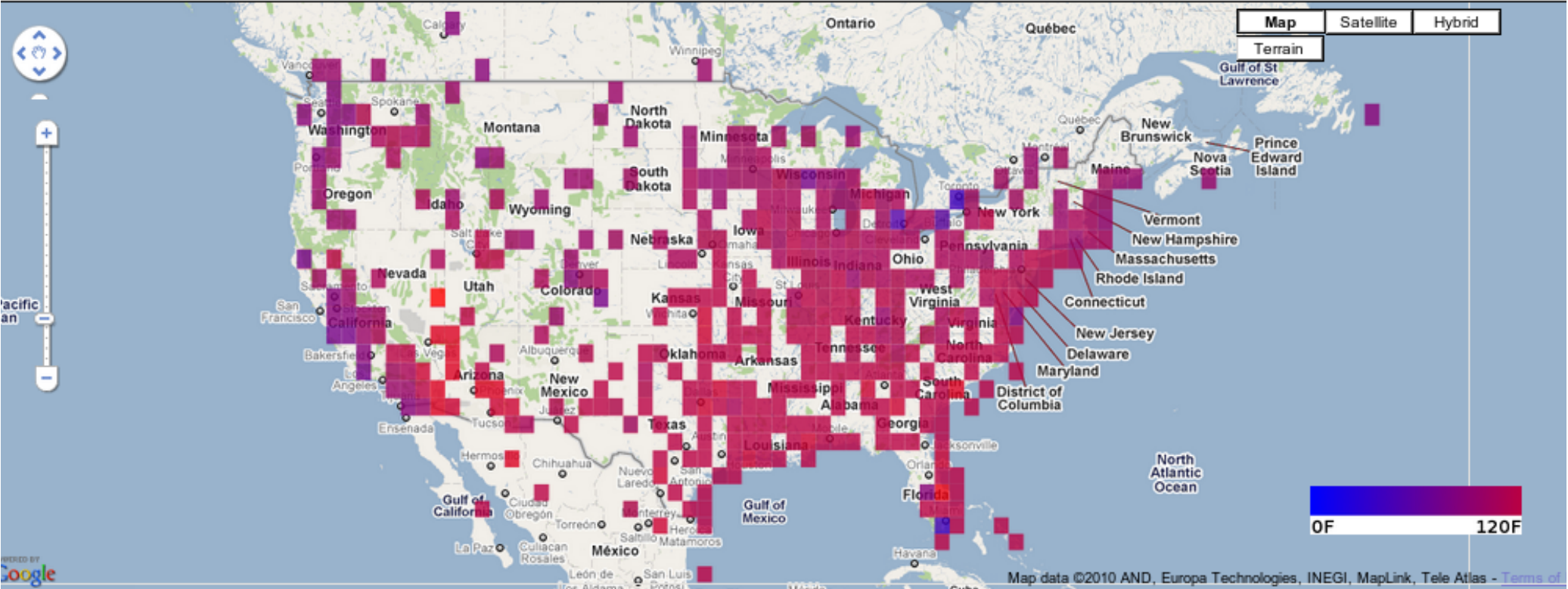
TweeQL demo  
\*nerd alert\*



# TweeQL's other features

- Aggregation
- Outlier detection
- Joins/mashups with other sources

Average temperatures extracted from tweets.

# Unstructured data

Unstructured  
data

+

Structured  
queries

Unstructured  
data

+

Structured  
queries

=> Structured  
data

Unstructured  
data

+

Structured  
queries

=> Structured  
data => Meaningful  
visualizations

# Thanks!

- TweepQL: <http://github.com/marcua/tweepql>
- TwitInfo: <http://twitinfo.csail.mit.edu/>
- Ask me about Mechanical Turk + Journalism!

Adam Marcus

@marcua

[marcua@csail.mit.edu](mailto:marcua@csail.mit.edu)

<http://people.csail.mit.edu/marcua>

## august 23 manchester city vs. liverpool

Keywords: football, soccer, epl, premier\_league, premierleague, manchester city, mancity, liverpool

Event dates: Aug. 23, 2010, 6:50 p.m. - Aug. 23, 2010, 9:10 p.m.

### Message Frequency



### Tweet Map



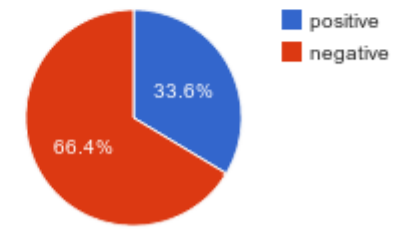
### Relevant Tweets

-  Tevez... Nets it man! ManCity 3-0 Liverpool... #mogbe
-  Man City 3-0 Liverpool (Tevez pen)
-  penalty for City. And Tevez makes it 3-0 for ManCity
-  Mancity vs livpool 3-0 tevez gooolllllll
-  BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66&#039;) <http://twurl.nl/knkx4l>

### Popular Links

- <http://bit.ly/aLlVUw> (cited by 5)
- <http://bit.ly/cxBF2H> (cited by 5)
- <http://twitcam.com/1rmka> (cited by 5)

### Overall Sentiment



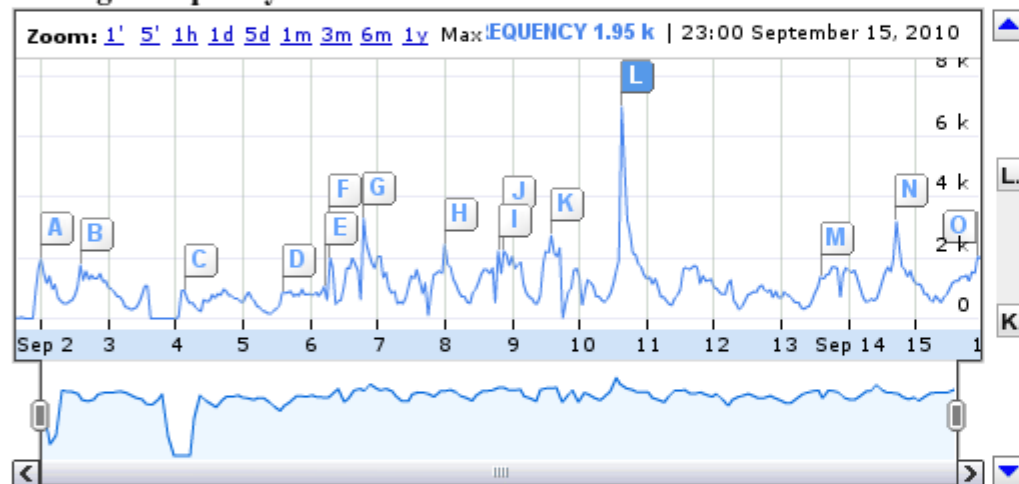


# barack obama

Keywords: obama

Event dates: Sept. 1, 2010, 3 p.m. - Sept. 16, 2010, midnight

## Message Frequency



## Tweet Map



## Relevant Tweets

**VOTE** Goolsbee Said to Be Picked to Head White House Economic Panel: President Barack Obama will appoint Austan Goolsbee... <http://bit.ly/birIPH>

**Top Adviser to Lead Panel on Economy - NYTimes.com:** President Obama on Friday will promote a longtime economic adv... <http://bit.ly/bklmF1>

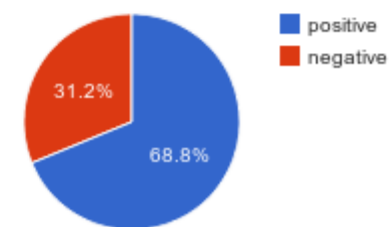
**WASHINGTON — President Barack Obama has chosen one of**

- administration [Label this event](#)
- 2010-9-13
- L. obama, president, economic, barack, economy** [Label this event](#)
- 2010-9-10
- K. obama, burning, koran, quran, burn** [Label this event](#)
- 2010-9-9

## Popular Links

- <http://bit.ly/czJdxP> (cited by 22)
- <http://lat.ms/dcig1f> (cited by 4)
- <http://bit.ly/9XzAyd> (cited by 4)

## Overall Sentiment

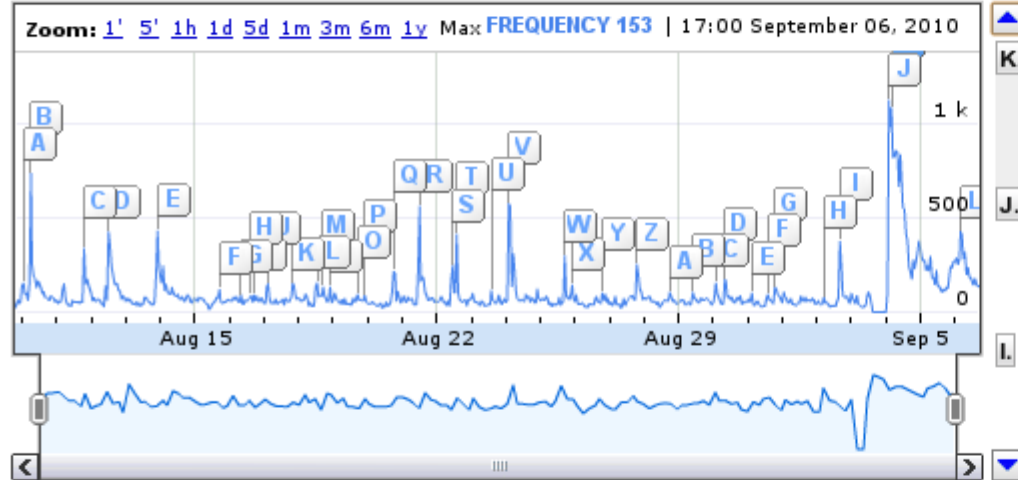


# earthquakes

Keywords: earthquake

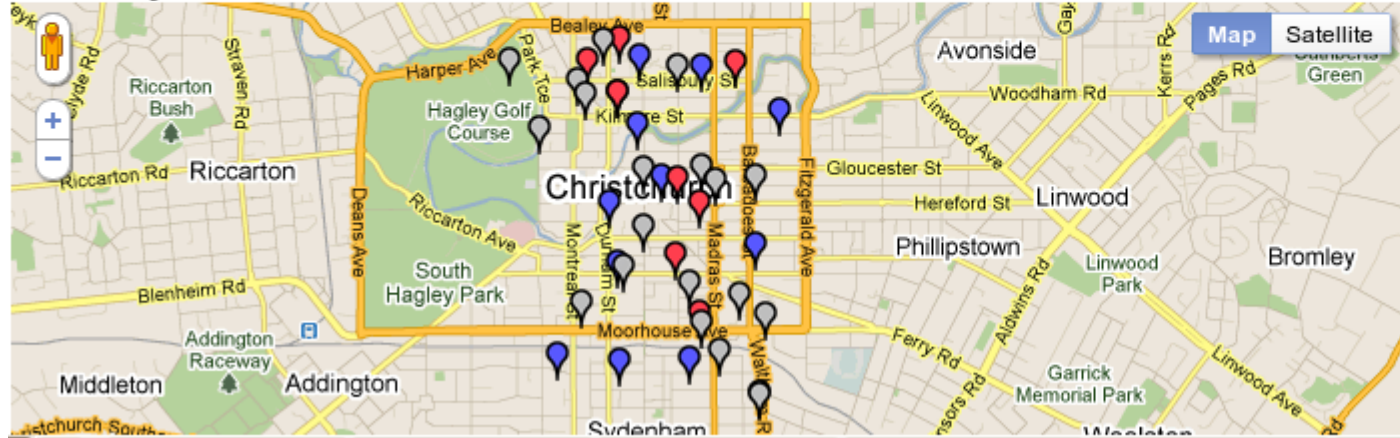
Event dates: Aug. 9, 2010, 7:18 p.m. - Sept. 6, 2010, 6 p.m.

## Message Frequency



- Enter a filter term
- K.** earthquake, christchurch, zealand, thoughts, prayers  
[Label this event](#)  
2010-9-4
  - J.** earthquake, christchurch, 7.0, zealand, nz  
[Label this event](#)  
2010-9-4
  - I.** earthquake, 3.3, felt, pablo, san  
[Label this event](#)  
2010-9-2

## Tweet Map



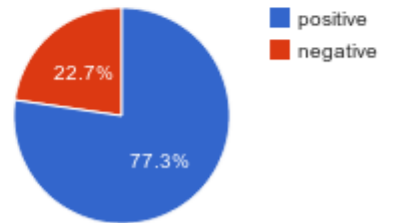
## Relevant Tweets

-  RT @mariacelina: This morning, my thoughts & prayers go out to those affected by the earthquake that recently hit Christchurch, New Zealand.
-  RT @KennyHamilton: Just heard about the earthquake in New Zealand...my thoughts and prayers go out to everyone in that country. I hope ...
-  RT @EvermoreBand: Thoughts and prayers with the people of Christchurch. Massive earthquake there, and no power or water for the whole city.

## Popular Links

- <http://on.cnn.com/8YDPyG> (cited by 11)
- <http://bit.ly/ddWgtK> (cited by 6)
- <http://bit.ly/aFh7RM> (cited by 5)

## Overall Sentiment



# TweeQL Lessons Learned

- Geographic limitations
- Requires grungy regular expressions
- Data is not necessarily relational

# Behind the Hunt for Bin Laden



Reuters

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. [More Photos »](#)

Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

Bin Laden Is Dead, Obama Says



An advertisement for Travel Dyer. It features a blue background with a white and orange logo at the top right that says "Travel Dyer Excellence in Service". Below the logo, the text reads "Explore the splendors of the all-suite, 114-room, 34-guest private villa" over a scenic view of a coastal town with red-roofed buildings and a blue sea.



# Behind the Hunt for Bin Laden



Reuters

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. [More Photos »](#)

Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

Bin Laden Is Dead, Obama Says



An advertisement for Travel Dyer. It features a logo with a stylized orange and blue wave above the text "Travel Dyer" and "Excellence in Service". Below the logo, the text reads "Explore the splendors of the all-suite, 114-room, 34-guest private villa" over a scenic background of a coastal town with red-roofed buildings and a blue sea.



# Behind the Hunt for Bin Laden



Reuters

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. [More Photos »](#)

Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

[Bin Laden Is Dead, Obama Says](#)



An advertisement for Travel Dye. It features a blue background with a white and orange logo at the top that says "Travel Dye" and "Excellence in Service". Below the logo, the text reads "Explore the splendors of the all-suite, 114-room, 34-guest private villa". The bottom of the ad shows a scenic view of a coastal town with red-roofed buildings built on a cliffside overlooking the sea.

1 | **2** | 3 | 4 | [NEXT PAGE »](#)

*Reporting was contributed by Elisabeth Bumiller, Charlie Savage and Steven Lee Myers from Washington, Adam Ellick from New York, and Ismail Khan from Peshawar, Pakistan.*



# Behind the Hunt for Bin Laden



Reuters

President Obama and members of his national security team receiving an update on Sunday. A classified document in front of Hillary Rodham Clinton was blurred before this photo was released. [More Photos »](#)

Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

[Bin Laden Is Dead, Obama Says](#)



An advertisement for Travel Dyer. It features a blue background with a white and orange logo at the top that says 'Travel Dyer Excellence in Service'. Below the logo, the text reads 'Explore the splendors of the all-suite, 114-room, 34-guest private villa'. The bottom half of the ad shows a scenic view of a coastal town with red-roofed buildings built on a cliffside overlooking the sea.

*Reporting was contributed by Elisabeth Bumiller, Charlie Savage and Steven Lee Myers from Washington, Adam Ellick from New York, and Ismail Khan from Peshawar, Pakistan.*

# Behind the Hunt for Bin Laden



Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

[Bin Laden Is Dead, Obama Says](#)




Travel Dynasty  
Excellence in Service

[Explore the splendors](#)


 Man City 3-0 Liverpool (Tevez pen)


 penalty for City. And Tevez makes it 3-0 for ManCity

 Mancity vs livpool 3-0 tevez gooolllllll

 BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66'&#039;) <http://twurl.nl/knkx41>

 3-0! Mancity vs liverpool

 gool de carlos tevez para el manchester city 3-0 liverpool

 Tevez scores and City now has a very commanding 3-0 lead over Liverpool



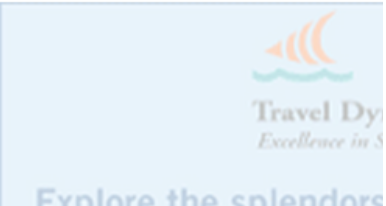
# Behind the Hunt for Bin Laden



Log in to see what your friends are sharing on nytimes.com. [Privacy Policy](#) | [What's This?](#)

## What's Popular Now

Bin Laden Is Dead, Obama Says



Man City 3-0 Liverpool (Tevez pen)



penalty for City. And Tevez makes it 3-0 for ManCity



Mancity vs livpool 3-0 tevez gooollllll

# Breadth



BOLANEWS.COM - Update: Manchester City - Liverpool 3-0 (Tevez 66'&#039;) <http://twurl.nl/knkx4l>




3-0! Mancity vs liverpool





gool de carlos tevez para el manchester city 3-0 liverpool



Tevez scores and City now has a very commanding 3-0 lead over Liverpool

 **markng** Mark Ng  
great way to start my morning. Kittens : Check ; #yes2av : Check -  
[http://www.youtube.com/watch?v=HiHuiDD\\_oTk](http://www.youtube.com/watch?v=HiHuiDD_oTk)  
2 minutes ago

 **newsyc20** Hacker News 20  
Browser tabs should indicate which pages are playing audio  
<http://news.ycombinator.com/item?id=2509624> #active  
3 minutes ago

 **PaulLewis** Paul Lewis  by markng  
Ian Tomlinson's reaction: verdict was "statement of the blindingly obvious"  
[gu.com/p/2zz5t/tw](http://gu.com/p/2zz5t/tw)  
10 minutes ago

 **derekcollison** Derek Collison  by antirez  
The #cfoundry community has been amazing. Thanks..  
<http://j.mp/iP6qea>  
7 minutes ago

 **markng** Mark Ng  
nice to see that the UK justice system seems to be working correctly around the Tomlinson police killing.  
7 minutes ago

 **newsyc20** Hacker News 20  
Getting Started with Android Development <http://bit.ly/mgS4Eo>  
(<http://bit.ly/mpnBZz>) via @razin #guru  
8 minutes ago

 **NiemanLab** Nieman Lab  
PBS just got more Google friendly thanks to transcripts of more

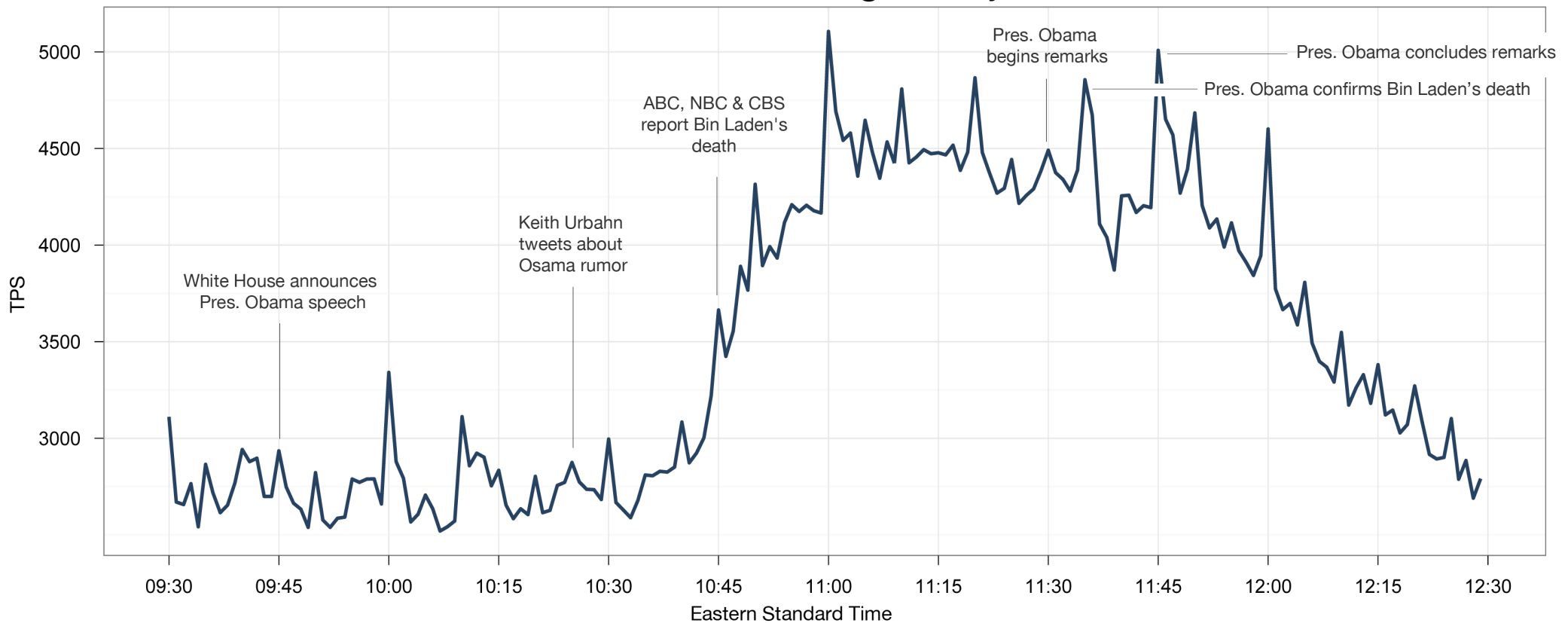
markng Mark Ng  
great way to start my morning. Kittens : Check ; #yes2av : Check -  
[http://www.youtube.com/watch?v=HiHuiDD\\_oTk](http://www.youtube.com/watch?v=HiHuiDD_oTk)  
2 minutes ago

newsyc20 Hacker News 20  
Browser tabs should indicate which pages are playing audio  
<http://news.ycombinator.com/item?id=2509624> #active  
3 minutes ago

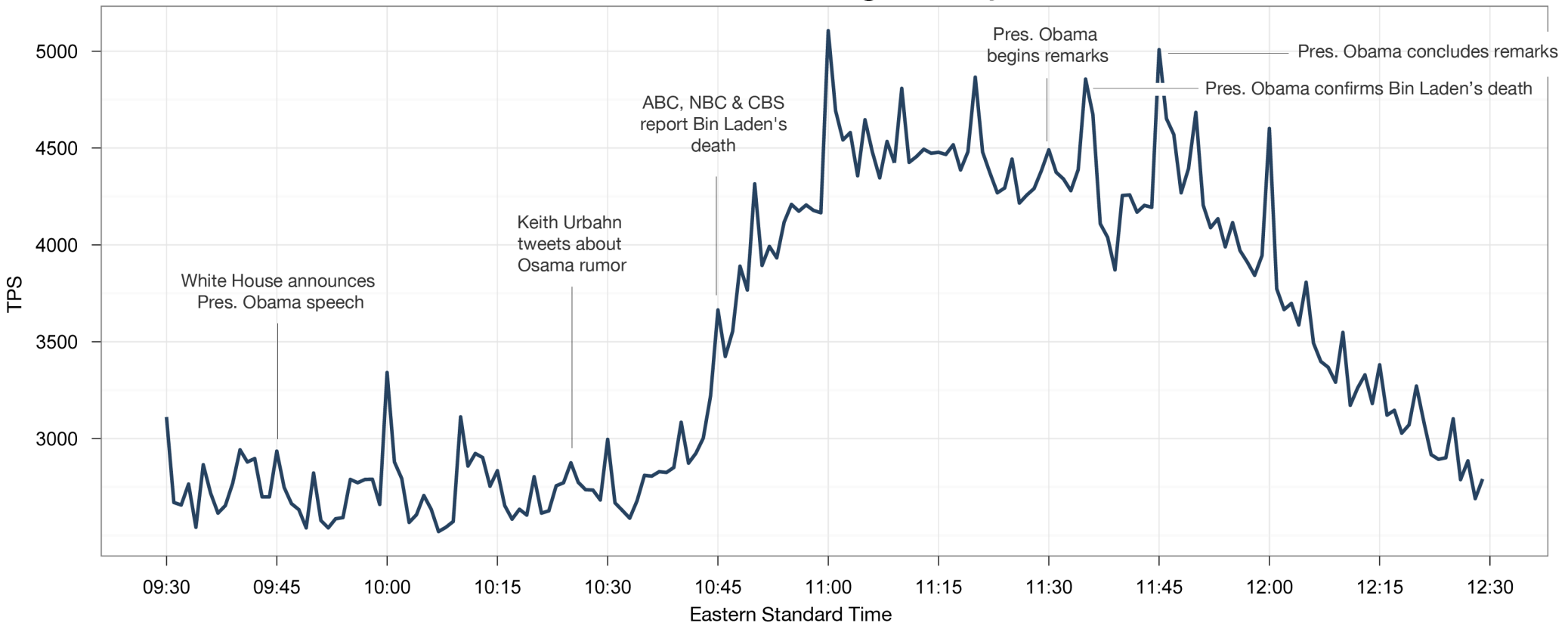
PaulLewis Paul Lewis by markng  
Ian Tomlinson's reaction: verdict was "statement of the blindingly obvious"  
[gu.com/p/2zz5t/tw](http://gu.com/p/2zz5t/tw)  
10 minutes ago

derekcollison Derek Collison by antirez  
The #cfoundry community has been amazing. Thanks..  
<http://j.mp/iP6qea>  
7 minutes ago

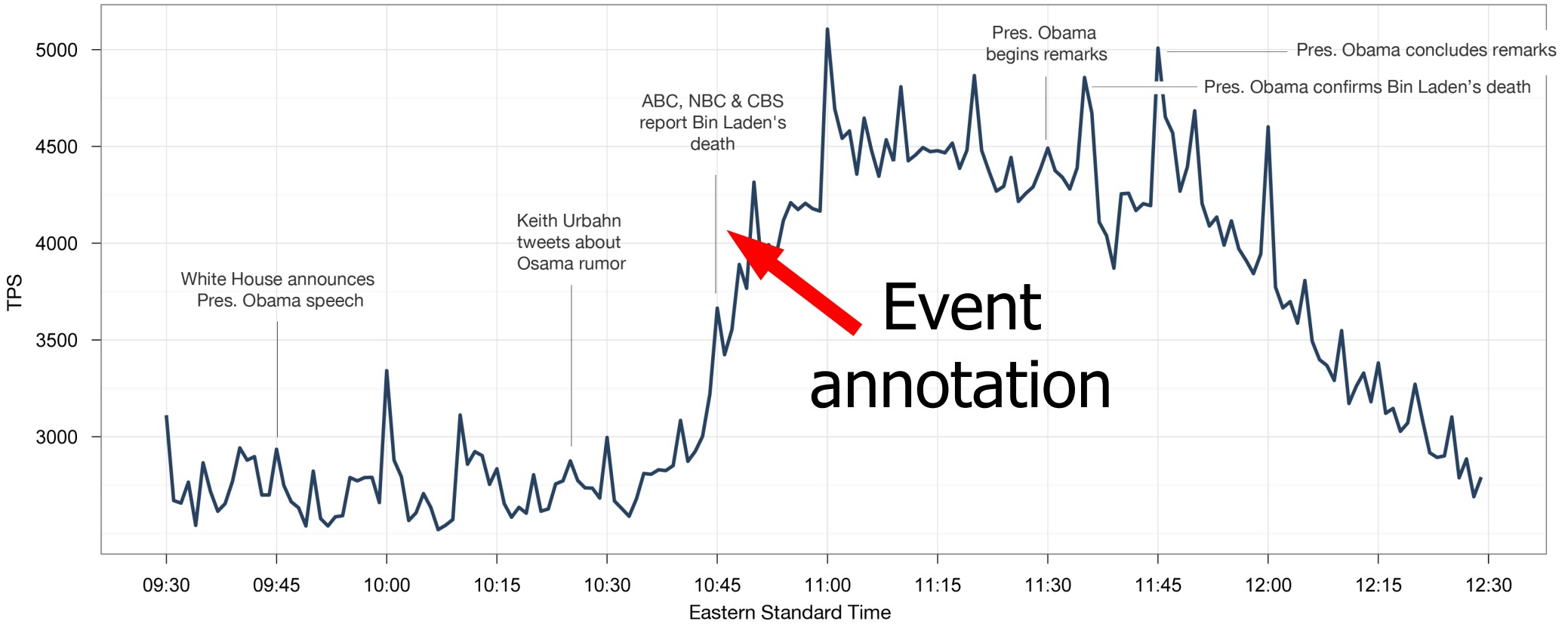
## Tweets Per Second: Evening of May 1, 2011



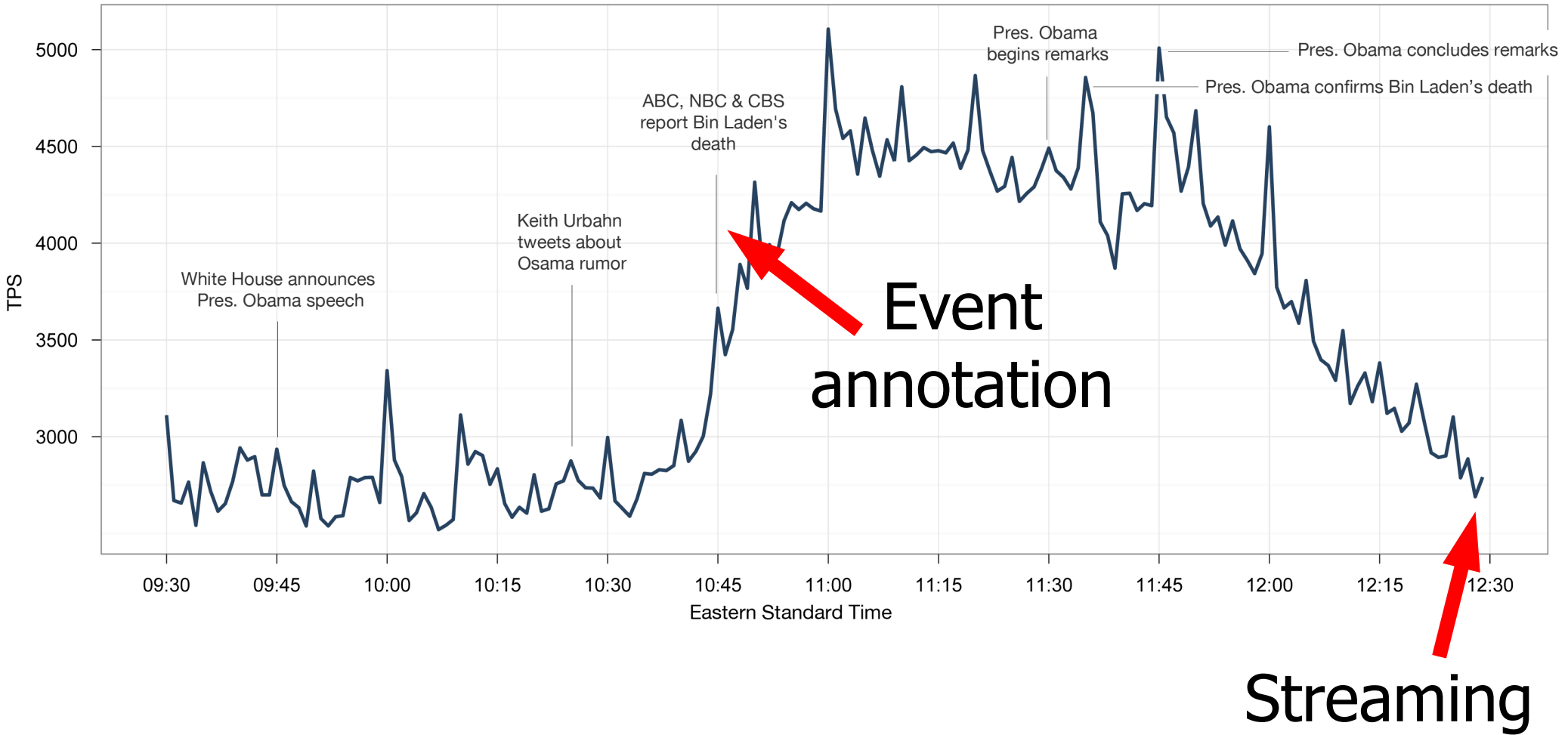
# Tweets Per Second: Evening of May 1, 2011



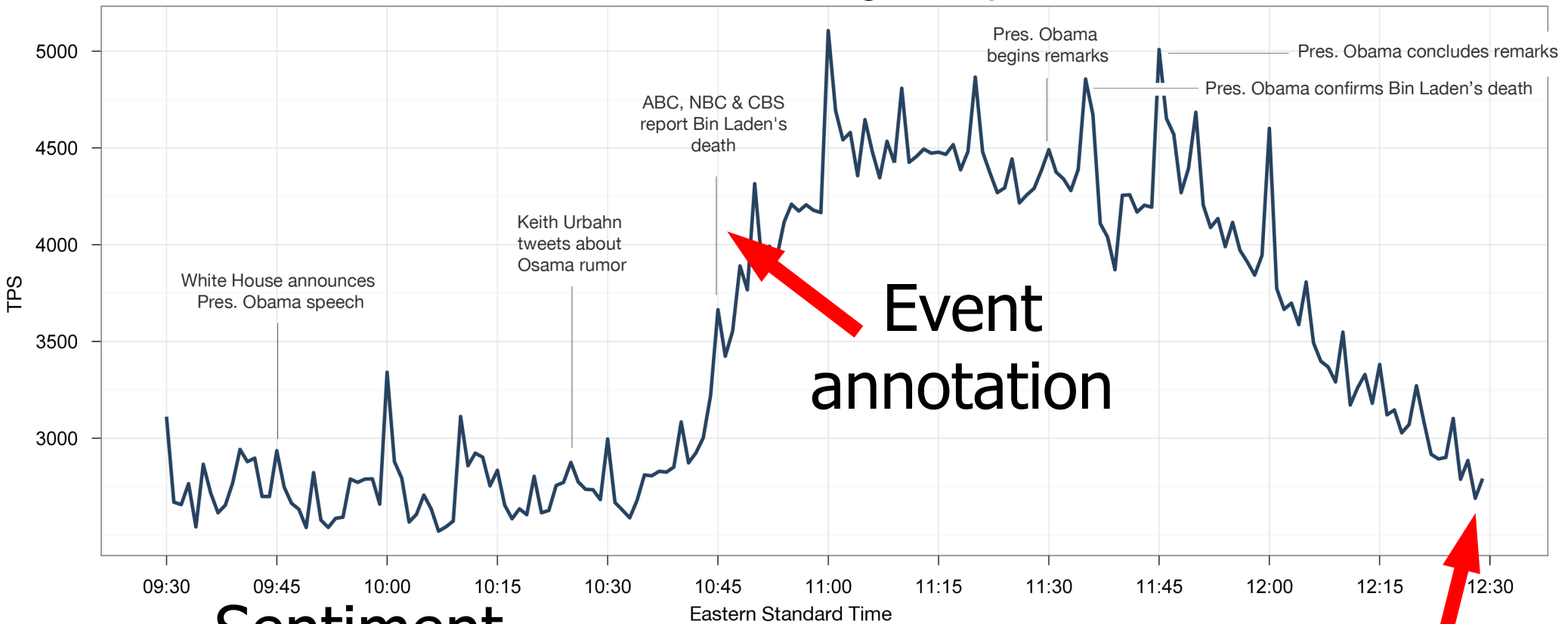
# Tweets Per Second: Evening of May 1, 2011



# Tweets Per Second: Evening of May 1, 2011



# Tweets Per Second: Evening of May 1, 2011



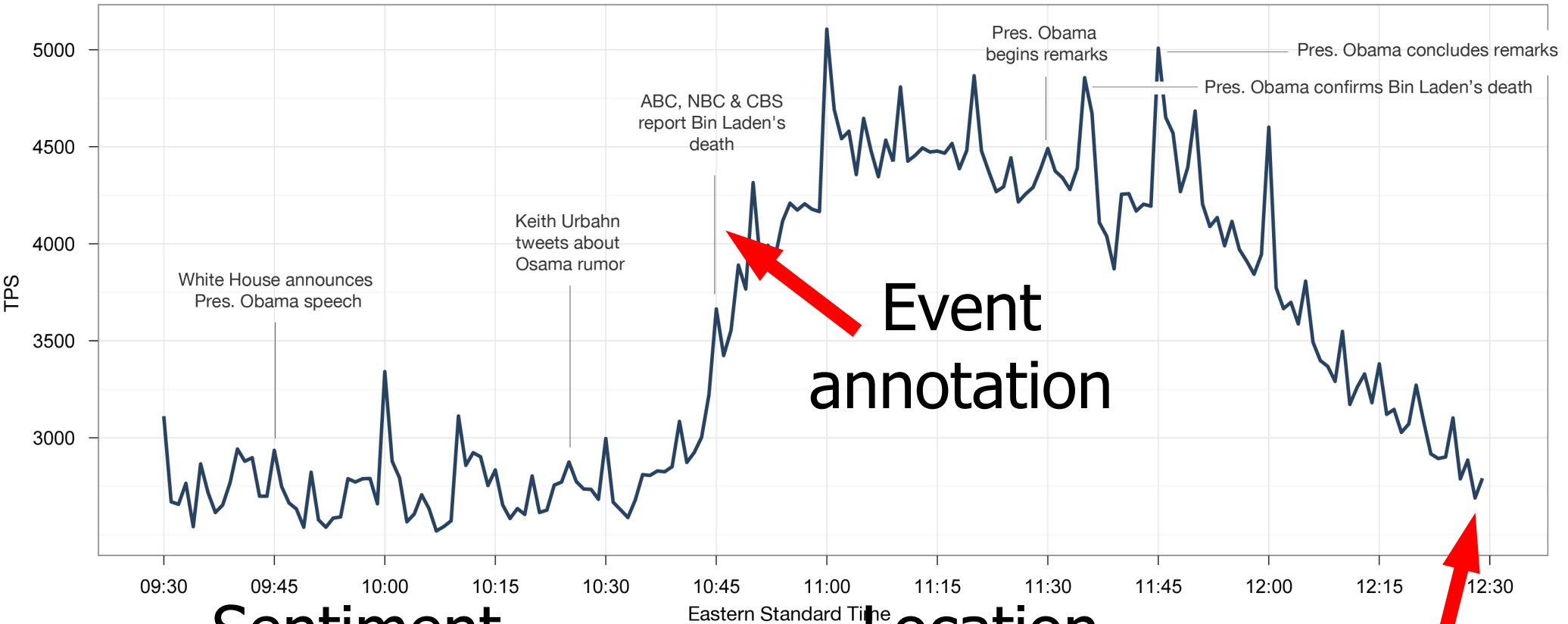
Sentiment



@laurenceb  
Laurence Berland

Again, glad UBL dead, but this jubilation troubles me. The dead are still dead. We ought strive for a world where killing isn't necessary.

# Tweets Per Second: Evening of May 1, 2011



Sentiment



@laurenceb  
Laurence Berland

Again, glad UBL dead, but this jubilation troubles me. The dead are still dead. We ought strive for a world where killing isn't necessary.

Location



@harrisj  
Jacob Harris

Do we have an idea how many tweets were from within the US vs. elsewhere last night? Otherwise twitter users <> US network compare flawed