



# AN EXAMINATION OF REGRET IN BULLYING TWEETS

**Jun-Ming Xu, Benjamin Burchfiel, Xiaojin Zhu**

Department of Computer Sciences

**Amy Bellmore**

Department of Educational Psychology

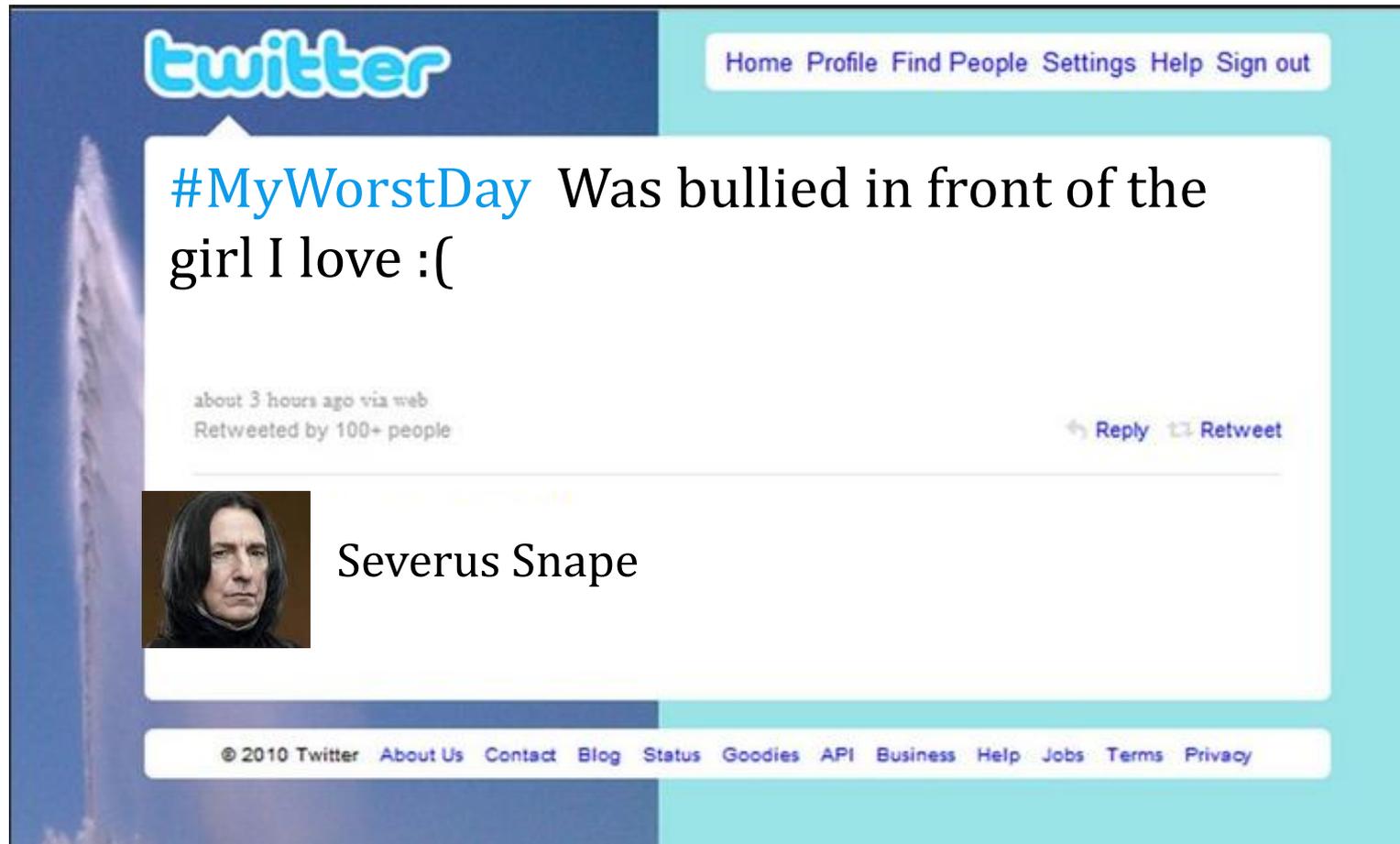
University of Wisconsin-Madison

# Bullying: Snape's Worst Memory



<http://harry-potter-spain.deviantart.com/art/Snape-s-Worst-Memory-27310861>

# If they were Twitter users...



The image shows a screenshot of a Twitter post. At the top left is the Twitter logo. To the right is a navigation bar with links: Home, Profile, Find People, Settings, Help, Sign out. The main content is a tweet with the text "#MyWorstDay Was bullied in front of the girl I love :(". Below the text, it says "about 3 hours ago via web" and "Retweeted by 100+ people". To the right of the tweet are "Reply" and "Retweet" buttons. Below the tweet is a profile picture of Severus Snape and his name "Severus Snape". At the bottom is a footer with copyright information and various links: © 2010 Twitter, About Us, Contact, Blog, Status, Goodies, API, Business, Help, Jobs, Terms, Privacy.

twitter

Home Profile Find People Settings Help Sign out

#MyWorstDay Was bullied in front of the girl I love :(

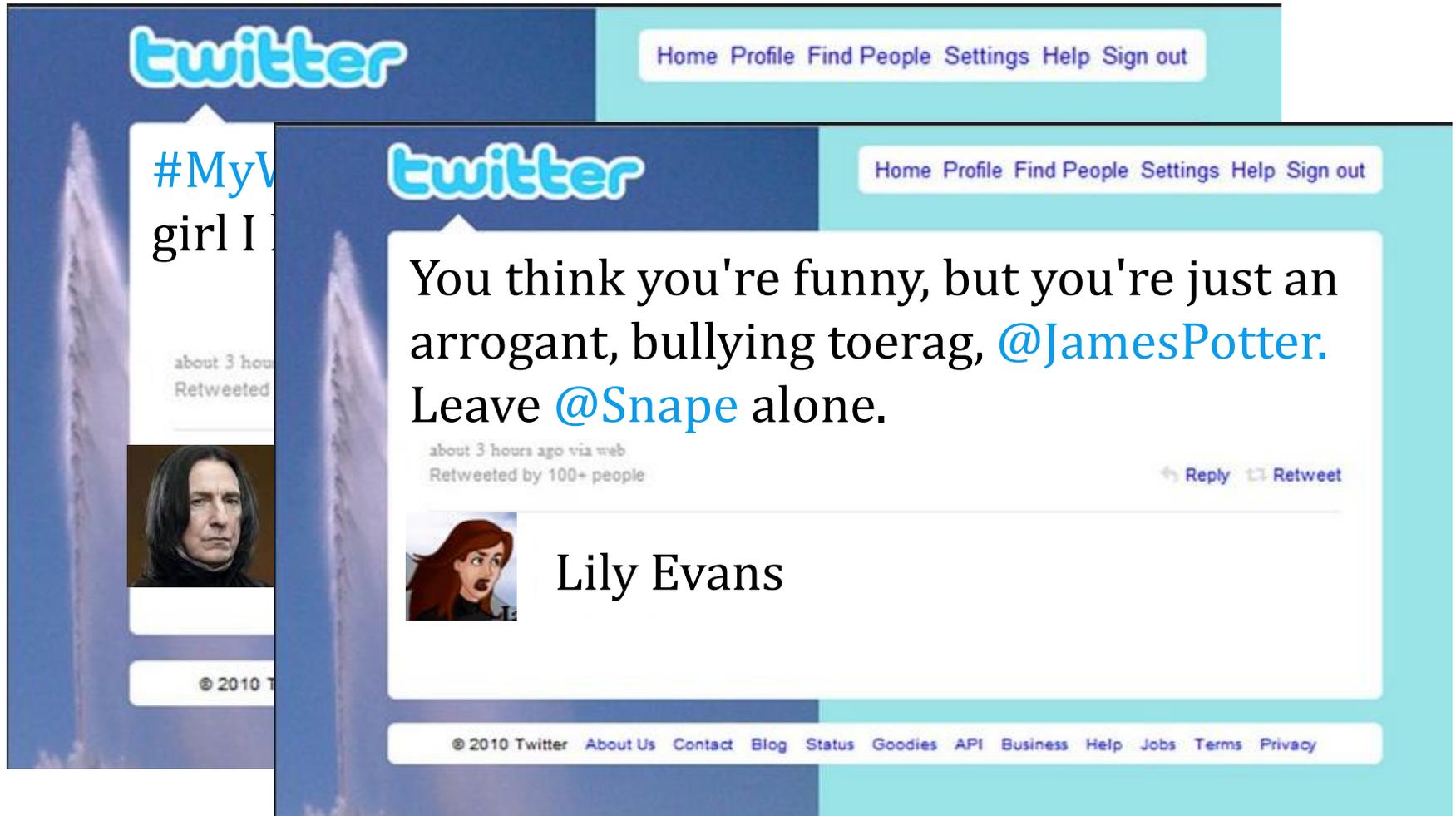
about 3 hours ago via web  
Retweeted by 100+ people

Reply Retweet

 Severus Snape

© 2010 Twitter About Us Contact Blog Status Goodies API Business Help Jobs Terms Privacy

# If they were Twitter users...



# Regret in Bullying Tweets

---

People post tweets they later regret

Potential risks

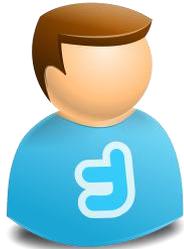
Embarrassment, Revictimization, Social ostracization

Our Goals

Isolate the factors associated with deleted tweets

Identify tweets that might be deleted later

# Data Collection



0:00

#MyWorstDay was **bullied** in front of the girl I love :(

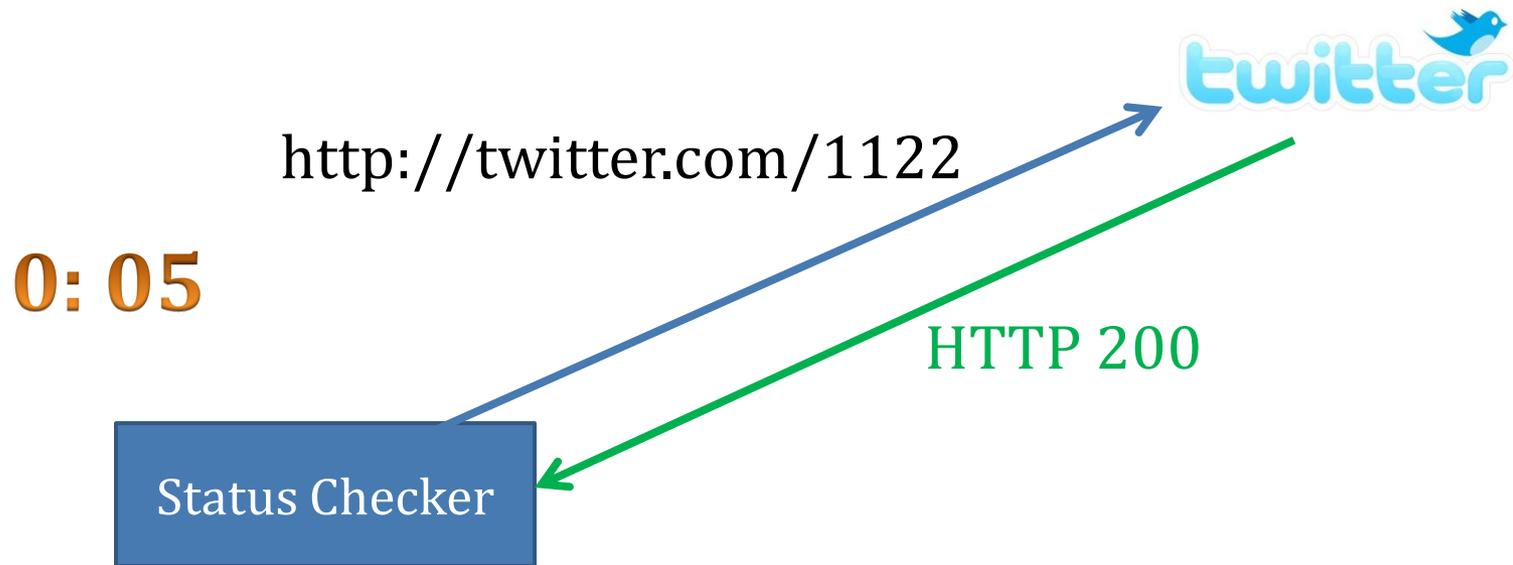


Status Checker

Bullying-or-not  
Text Classifier  
[Xu *et al.*, 2012]

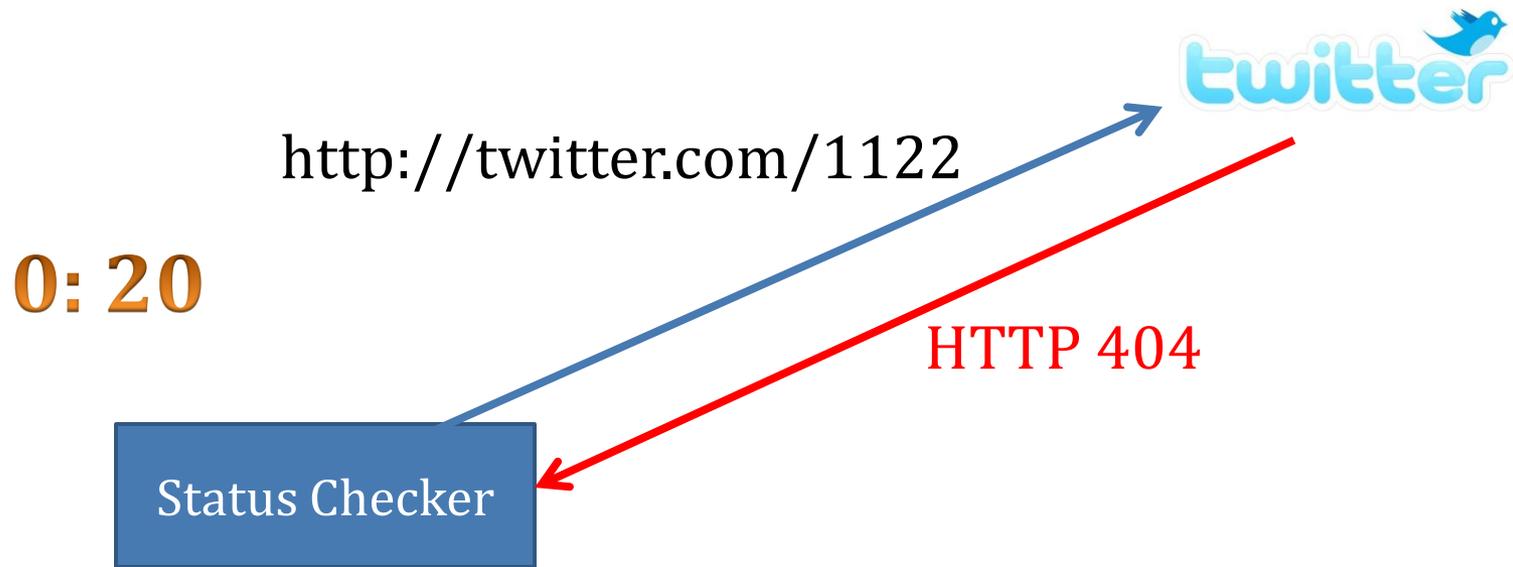
Tweet ID	Creation Time	1 <sup>st</sup> Check Time	2 <sup>nd</sup> Check Time	3 <sup>rd</sup> Check Time	...
1122	0:00	0:05	0:20	1:20	...

# Data Collection



Tweet ID	Creation Time	1 <sup>st</sup> Check Time	2 <sup>nd</sup> Check Time	3 <sup>rd</sup> Check Time	...
1122	0:00	0:05 ✓	0:20	1:20	...

# Data Collection



Tweet ID	Creation Time	1 <sup>st</sup> Check Time	2 <sup>nd</sup> Check Time	3 <sup>rd</sup> Check Time	...
1122	0:00	0:05 ✓	0:20 ✗	1:20	...

# Data Collection (Cont.)

Tweet ID	5 mins	20 mins	1.5 hours	6 hours	1 day	4 days	2 weeks	Status
1	✓	✓	✓	✓	✓	✓	✓	Surviving
2	✓	✓	?	?	?	✓	✓	Surviving
3	✓	✗	✗	?	?	?	?	Deleted
4	✓	✓	✓	✓	✗	✗	?	Deleted
...								

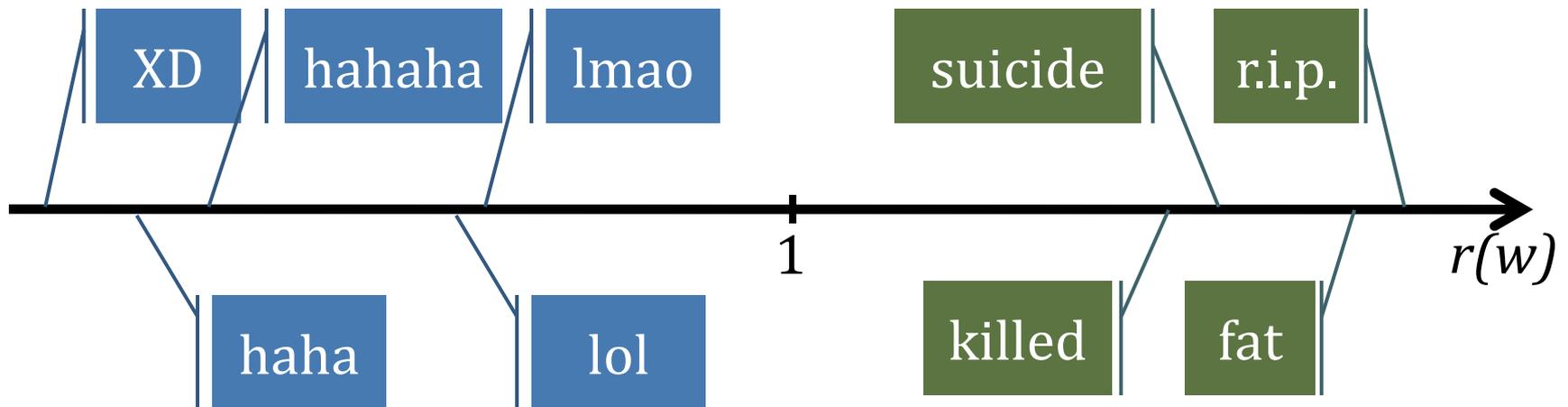
August - October, 2012

311,237 bullying tweets

# Joking words appear less frequently in deleted tweets

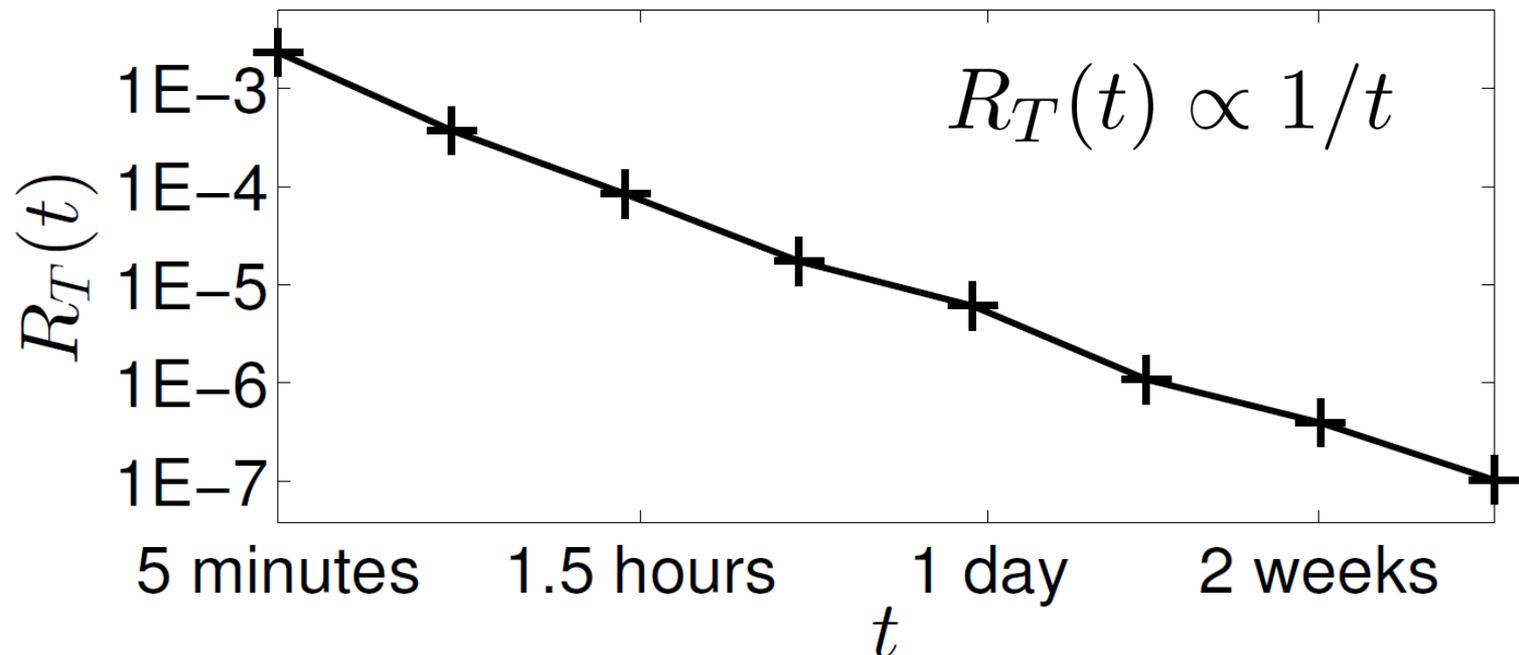
Odds ratio of a word  $w$

$$r(w) = \frac{P(w \mid \text{deleted})}{P(w \mid \text{surviving})}$$



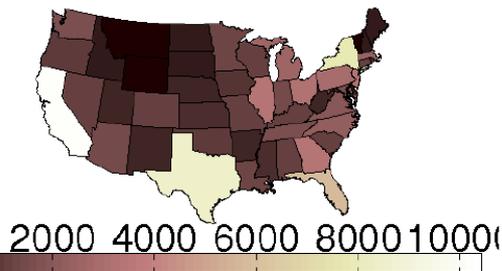
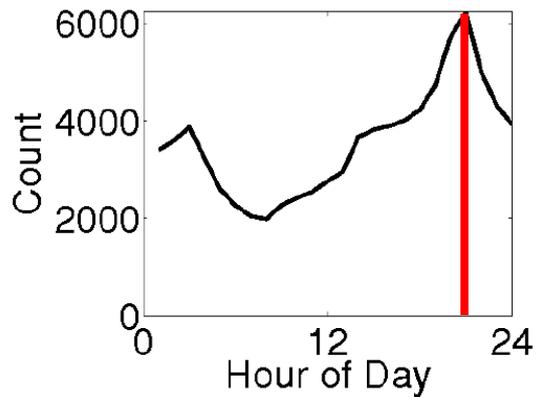
# Younger tweets more likely to be deleted

$R_T(t)$  is the fraction of tweets that are deleted during the one minute period  $(t, t + 1)$  after creation

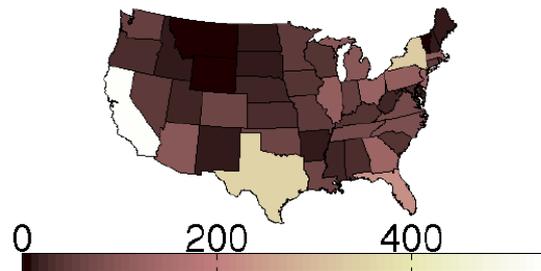
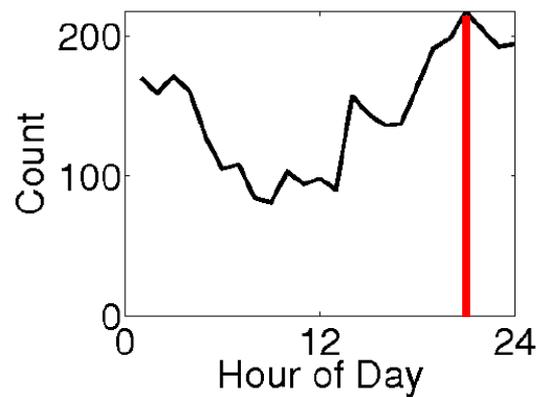


# Tweets created at late night more likely to be deleted

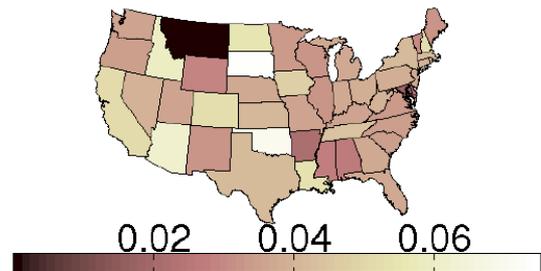
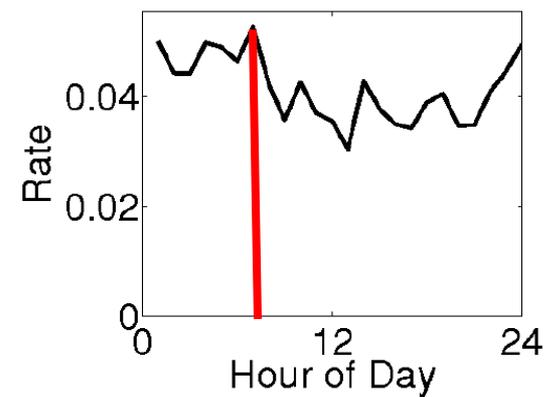
#Bullying tweets



#Deleted Bullying tweets



Deletion Rate



# Tweets authored by Reporter or Victim more likely to be deleted

Role	Deleted	Total	P(Deleted   Role)
Bully	1792	30123	5.95%
Victim	6497	83412	7.79%
Accuser	2541	50088	5.07%
Reporter	11370	147164	7.73%

# Teasing tweets less likely to be deleted

*“I may bully you but I love you lots. Just like jelly tots!”*

*“@USERNAME lol stop being a cyber bully lol :p.”*

Teasing?	Deleted	Total	P(Deleted   Teasing?)
Yes	858	22876	3.75%
No	21383	288361	7.42%

# Predict Regrettable Tweets

Subsample surviving tweets to balance classes  
22,241 deleted, 22,241 surviving

Text categorization

Case-folding, anonymization, tokenization

Unigram + Bigram features, Linear SVMs

Cross validation accuracy  $0.607 \pm 0.012$

Baseline (majority vote) 0.5

# Conclusions

- Many factors are associated with deleted posts  
word usage, surviving time, creation time, author's role, teasing
- It's hard to predict deletion
- Context information might be important  
recent posts, historical behavior, social network

# Thanks!

<http://research.cs.wisc.edu/bullying>

[Bullying Map](#)

We thank Kwang-Sung Jun, Angie Calvin, and Charles Dyer for helpful discussions.

This work is supported by National Science Foundation grants IIS-1216758 and IIS-1148012.