

# Internet Application Censorship: Studies of Weibo in China and Twitter in Turkey

Dan S. Wallach, *Rice University*

# Today, two short studies

*Weibo*: at one point China's largest social network.

*Twitter*: broadly used worldwide, particularly popular in the Middle East.

*We're looking at web applications, not at networks or the "Great Firewall of China".*

# The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions

Tao Zhu, *Independent Researcher*

David Phipps, *Bowdoin College*

Adam Pridgen, *Rice University*

Jedidiah R. Crandall, *University of New Mexico*

Dan S. Wallach, *Rice University*

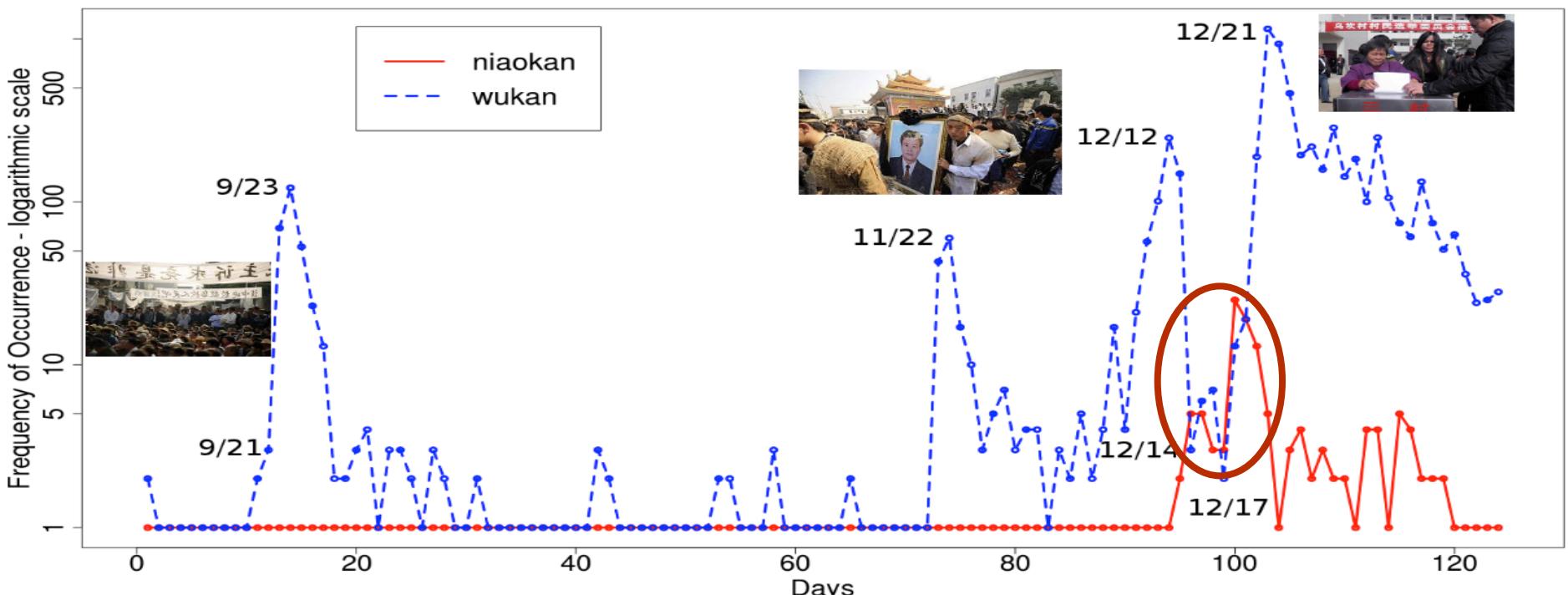
# Microblogging sites in China



# Sina Weibo

- 503 million registered users as of Dec 2012.
  - More than half are from mobile devices.
- About 100 million messages are posted each day on Sina Weibo.
- Promote visibility of social issues.

# Weibo's influence: Wukan incident - 2011



鸟坎 (The village name) vs 鸟坎 (Neologism)

# Sina Weibo

- Strict controls over the posts.

**cbcnews | Community**

Home World Canada Politics Business Health Arts & Entertainment

Community > Community Blog

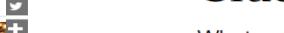
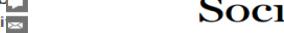
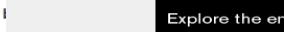
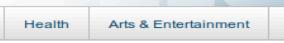
## Your Community Blog

**Chinese users skirt Weibo cen  
anniversary**

by John Bowman Posted: June 4, 2013 3:28 PM | Last Updated:  
Categories: Community, World

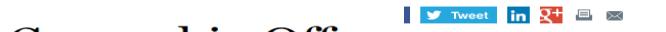
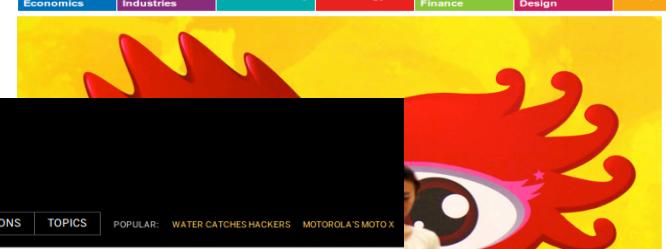
Tens of thousands of people gathered in Hong Kong to commemorate the 24th anniversary of the violent pro-democracy protesters in Tiananmen Square, I of the ceremony — or the calendar date at all — w from the social media site Sina Weibo.

These photos of the [candlelight vigil held in Hong Park](#) appeared on Weibo for just a few moments b deleted. The former British colony has been the si ceremonies since the 1989 massacre.



**Bloomberg Businessweek  
Global Economics**

Global Economics Companies & Industries Politics & Policy Technology Markets & Finance Innovation & Design Lifestyle



WEB NEWS

56 COMMENTS



## Social Media Censorship Offers Clues to China's Plans

What gets removed from China's social networks shows how censorship strategies are advancing, and can even hint at the government's plans.

By Tom Simonite on April 29, 2013

veen Sina Weibo ([SINA](#)) and the it allows the provocative to exist at all is that it must able of paid censors to scrub sive. A new academic study tials on how this process works.

researchers at Rice University, rsity of Mexico, and the or Tao Zhu wrote, "5% of the deletions [of controversial posts] happened in the first 8

# Introduction of our research

- Detecting a censorship event within 1-2 minutes of its occurrence.
- Three strategies Weibo system uses to target sensitive content quickly.

# Methodology

1. Identifying the sensitive user group
2. Crawling posts of sensitive user groups
3. Detecting deletions

# Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times



中国数字时代  
CHINA DIGITAL TIMES

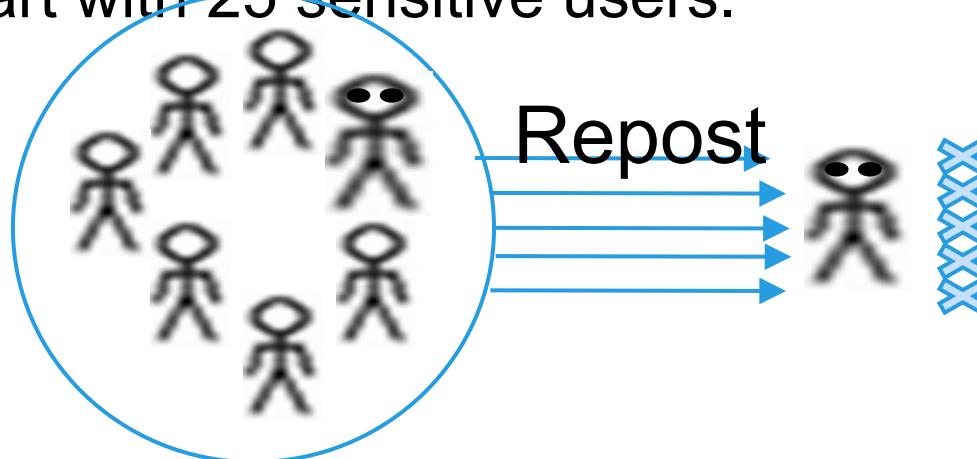
f t g+ 

敏感词	Sensitive Word	CDT English Post	第一次测试日期/结果 Date/Result of Test 1	第二次测试日期/结果 Date/Result of Test 2	第三次测试日期/结果 Date/Result of Test 3	第四次测试日期 Date/Result of
薄+刘	Bo+Liu	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄+死	Bo+death	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄+保外就医	Bo+medical parole	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄+(周)永康	Bo+(Zhou) Yongkang	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄王	Bo Wang	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄胡	Bo Hu	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
五个重庆	Five Chongqings	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
党内斗争	internal Party struggle	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
挺薄	support Bo	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
薄公	Master Bo	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
怀念+薄	think fondly of+Bo	<a href="http://goo.gl/xxKoSY">http://goo.gl/xxKoSY</a>	2013.7.25	1		
南华(早报)+马云	South China (Morning Post)+Jack Ma	<a href="http://goo.gl/mm0Ja">http://goo.gl/mm0Ja</a>	2013.7.20	1		

# Identifying the sensitive user group

- Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times.
- Start with 25 sensitive users.



# Identifying the sensitive user group

- Identifying the sensitive user group

- Use outdated sensitive keywords from China Digital Times.
- Start with 25 sensitive users.
- Sensitive group reaches 3,567 users after 15 days.
- More than 4,500 deletion daily

# Crawling

- User timeline:

- Weibo user timeline API returns the most recent 50 posts of the specified user.
- Query 3,567 sensitive users once per minute
  - 100 accounts for API call
  - 300 concurrent Tor circuits.
- Four-node cluster running Hadoop and Hbase
  - 2.38 million posts from July 20 to September 8, 2012.

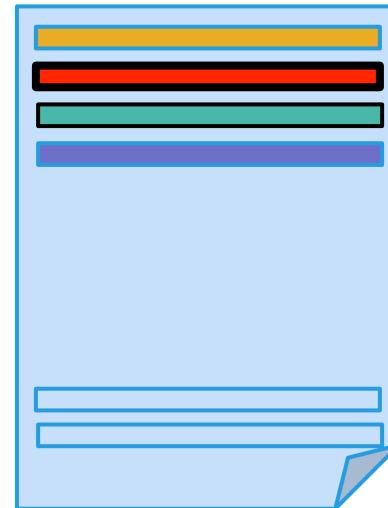
# Detecting deletions

Latest 50 posts



Diff

Our database



**Deleted Post**



# Detecting deletions

The lifetime of deleted Post =  $t_n - t_0$

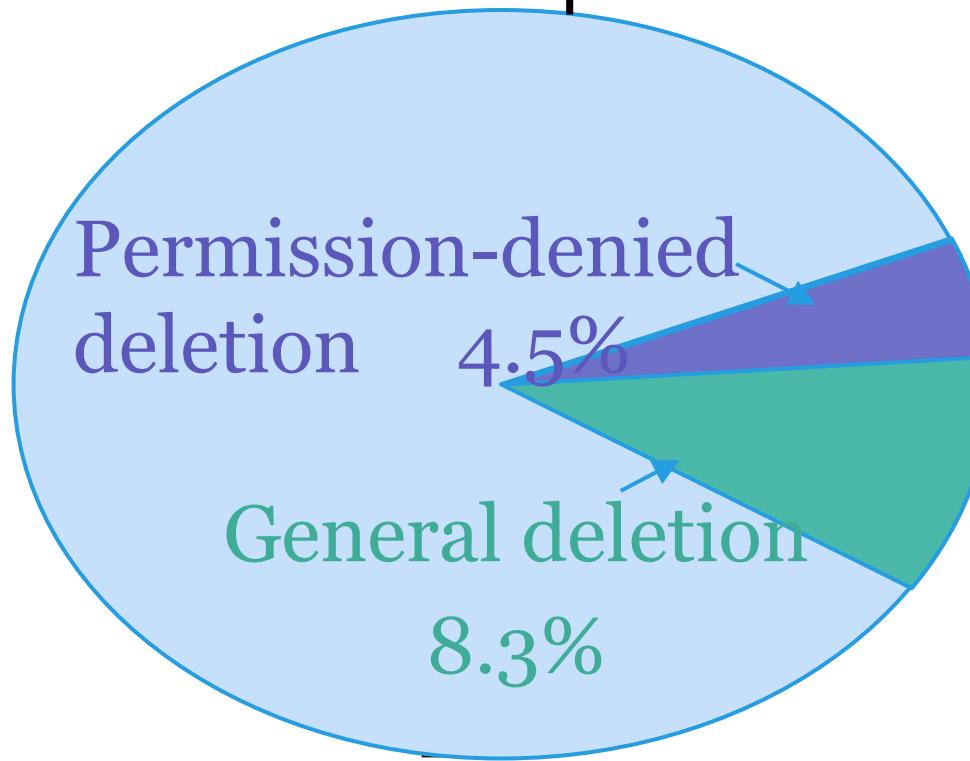


# Detecting deletions

- Permission-denied or system deletion
  - “Permission denied” error.
  - Caused by censorship events.
  - The post still exists but cannot be accessed by users.
  
- General deletion
  - “Post does not exist” error.
  - May caused by user self deletion or censorship events.
  - The post does not exist.

# Detecting deletions

2.38 Million user timeline posts



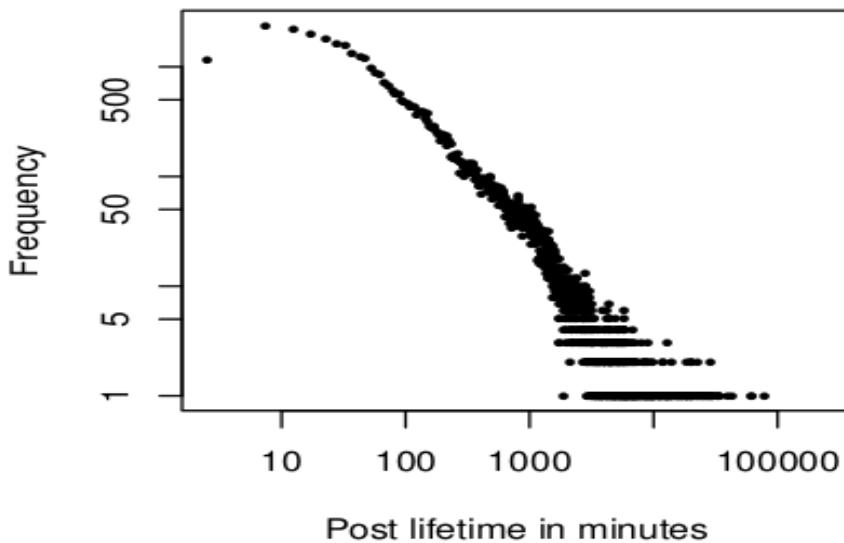
# Detecting deletions

- Permission-denied deletion or system deletion

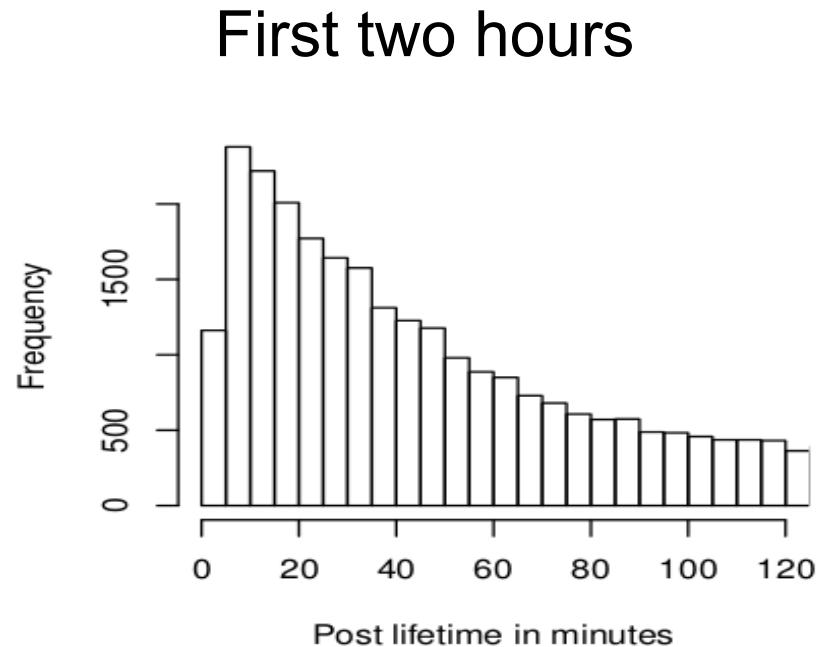
- Around 1,500 permission denied deletions.
- Comparing with WeiboScope, which is tracking around 300,000 users and have no more than 100 permission denied deletions daily.

# Distribution of deleted posts

Whole lifetime



First two hours



# Strategies to target sensitive contents

1. Weibo has filtering mechanisms as a proactive, automated defense.
2. Weibo targets specific users, such as those who frequently post sensitive content.
3. When a sensitive post is found, a moderator will use automated searching tools to find all of its related reposts, and delete them all at once.

# 1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive automated defense

Sorry, The content violates the relevant laws and regulations. If need help, please contact customer service.

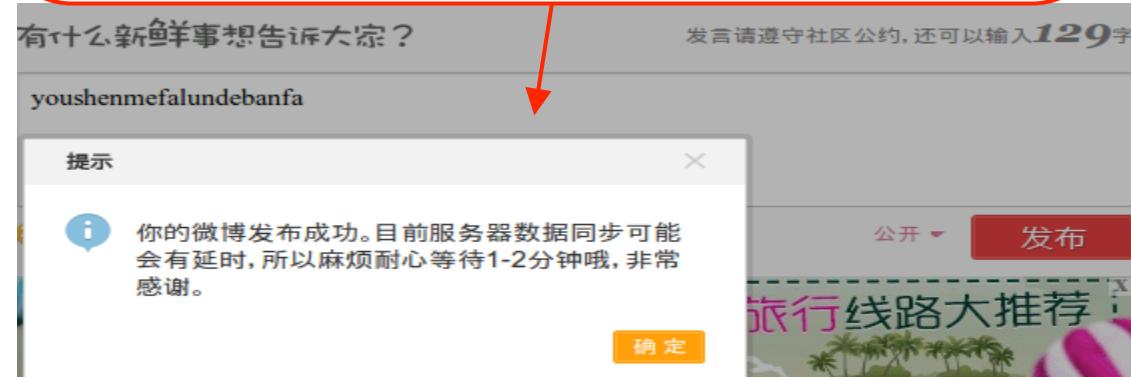


# 1. Keywords list filtering

- Weibo has filtering, proactive, automatic

- Explicit filtering
- Implicit filtering

Your post has been submitted successfully. Currently, there is a delay caused by server data synchronization. Please wait for 1 to 2 minutes. Thank you very much.

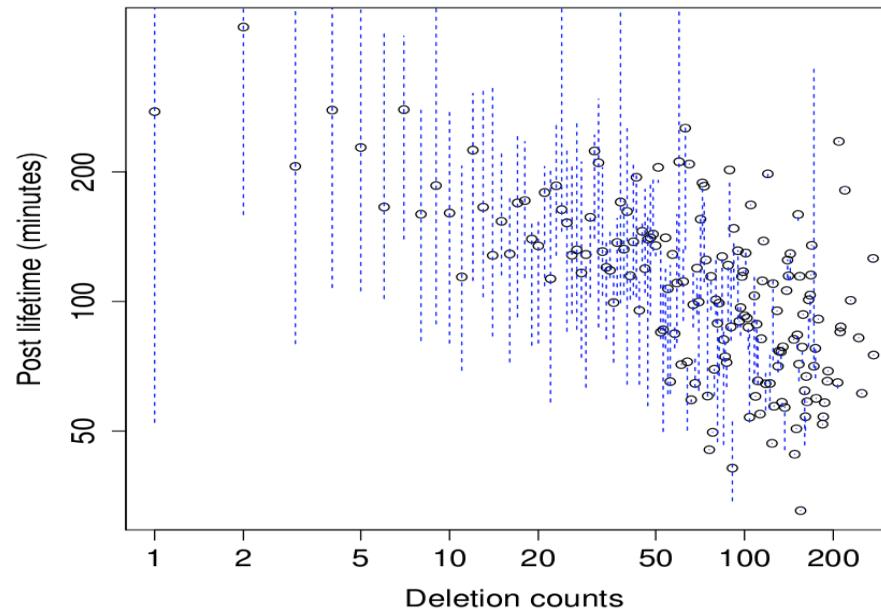


# 1. Keywords list filtering

- Weibo has filtering mechanisms as a proactive, automated defense
  - Explicit filtering
  - Implicit filtering
  - Camouflaged posts
  - Surveillance keywords list?
    - If no such list the cost will be too expansive

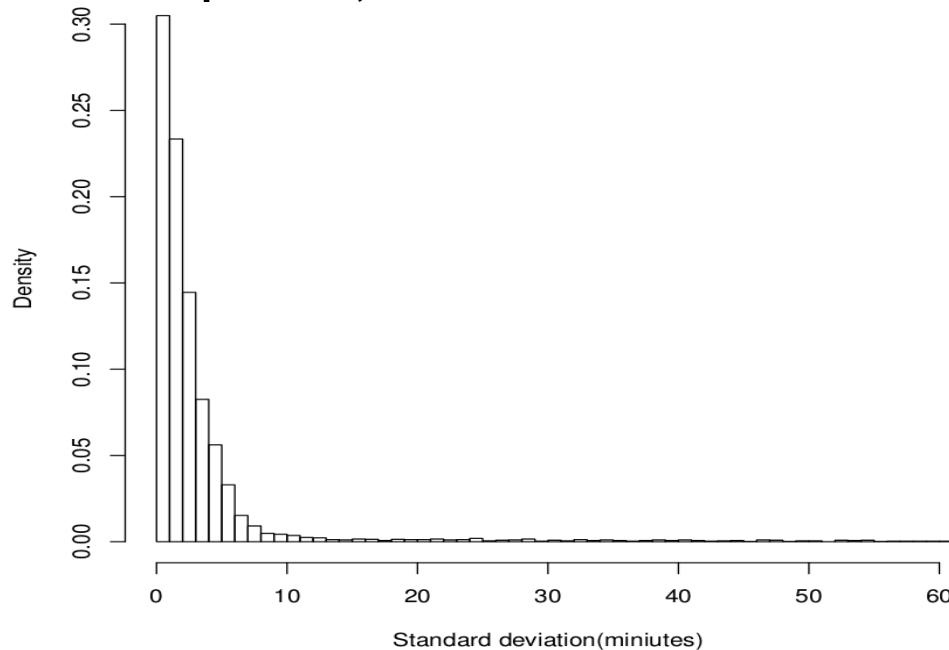
## 2. Targeting specific users

- Weibo targets specific users, such as those who frequently post sensitive content.

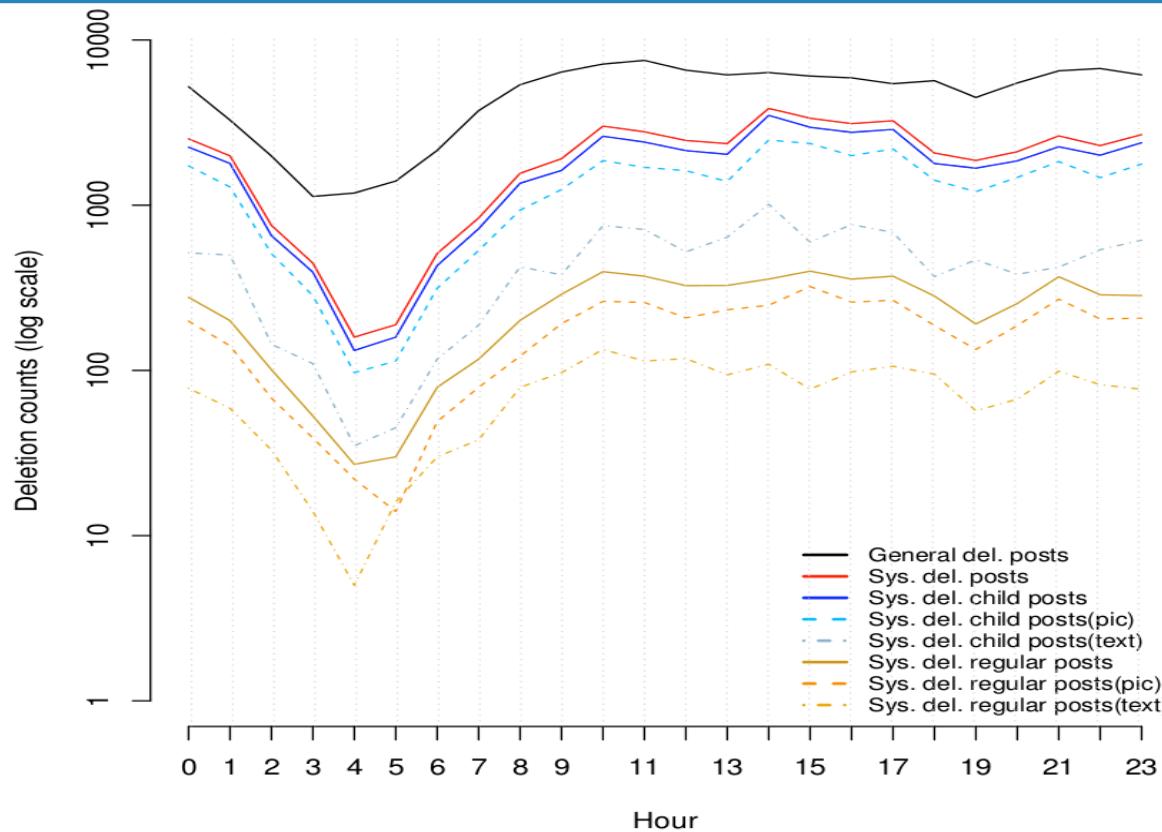


### 3. Finding all related reposts

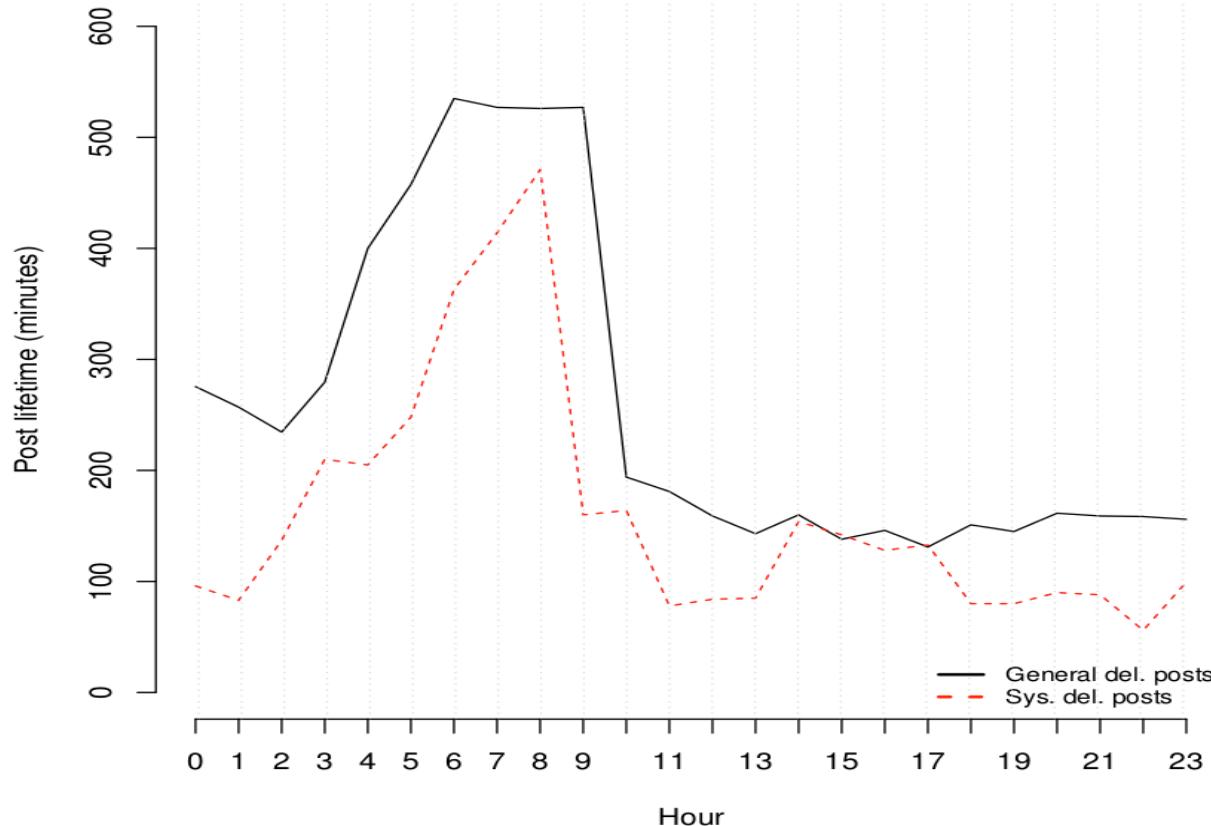
- When a sensitive post is found, a moderator can find all of its related reposts, and delete them all at once



# Censors work in the night

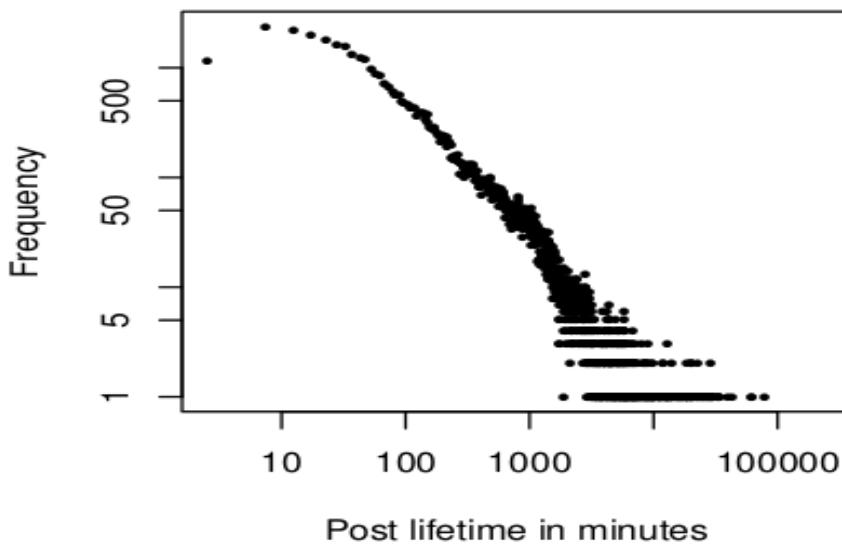


# Censors catch up in the morning

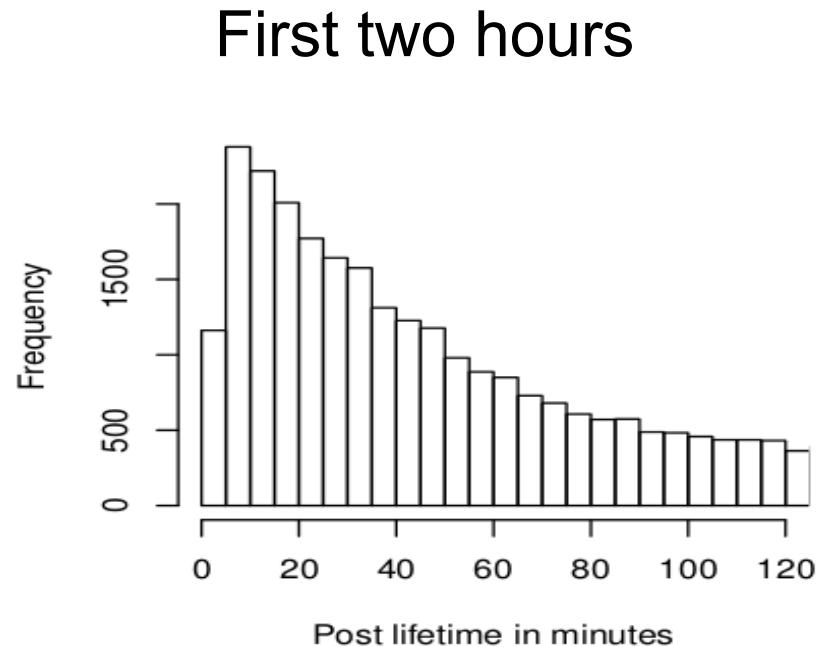


# Conclusion

Whole lifetime

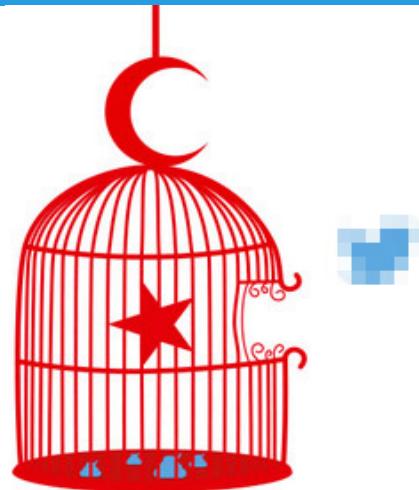


First two hours



# Known Unknowns: An Analysis of Twitter Censorship in #Turkey

Rima Tanash, *Rice University*  
Zhouhan Chen, *Rice University*  
Tanmay Thakur, *University of Houston*  
Chris Bronk, *University of Houston*  
Devika Subramanian, *Rice University*  
Dan S. Wallach, *Rice University*



# Disclaimer

- Our research is focused on quantifying the extent of Twitter censorship in a given country.
- We do not disclose private user information, or share Twitter datasets.

# Motivation

- Social media played a significant role during the recent wave of uprisings in the Middle East (a.k.a. Arab Spring)
  - Demonstrations were orchestrated via Twitter & Facebook.

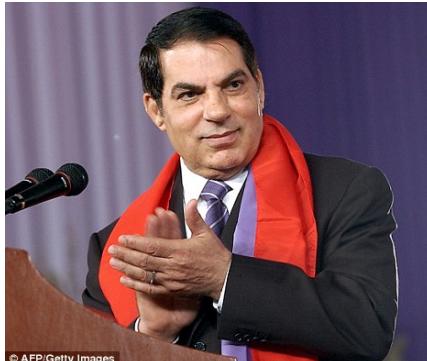


# Motivation

- Social media played a significant role during the recent wave of uprisings in the Middle East (a.k.a. Arab Spring)
  - Demonstrations were orchestrated via Twitter & Facebook.
  - Resulted in the overthrow of some Arab Dictatorships.

# Motivation

- Social media played a significant role during the recent wave of uprisings in the Middle East (a.k.a. Arab Spring)
  - Demonstrations were orchestrated via Twitter & Facebook.
  - Resulted in the overthrow of some Arab Dictatorships.



© AFP/Getty Images

# Government censorship

- Clearly, governments have great interest in controlling social media.

# Government censorship in the news

In Tunis  
protests

Netizen  
Social

Region

HOME

ACTIVITY WALL

WORLD

## EXCLUSIVE: Egypt Begins Surveillance Of Facebook, Twitter, And Skype On Unprecedented Scale

"We are looking at any conversation, any interaction, we might find worrying or would want to keep a closer eye on." Sheera Frenkel reports from Jerusalem and Maged Atef from Cairo for BuzzFeed News.

posted on Sept. 17, 2014, at 11:20 a.m.



# Types of censorship

- **Network level:**
  - Block entire service.
  - China blocks western social media.
  - Turkey blocks Twitter March, 2014.
- **Application level:**
  - Service is not blocked, censorship is internal to the service
  - Example: *Sina Weibo* (Zhu et. al) – keyword filtering, etc.

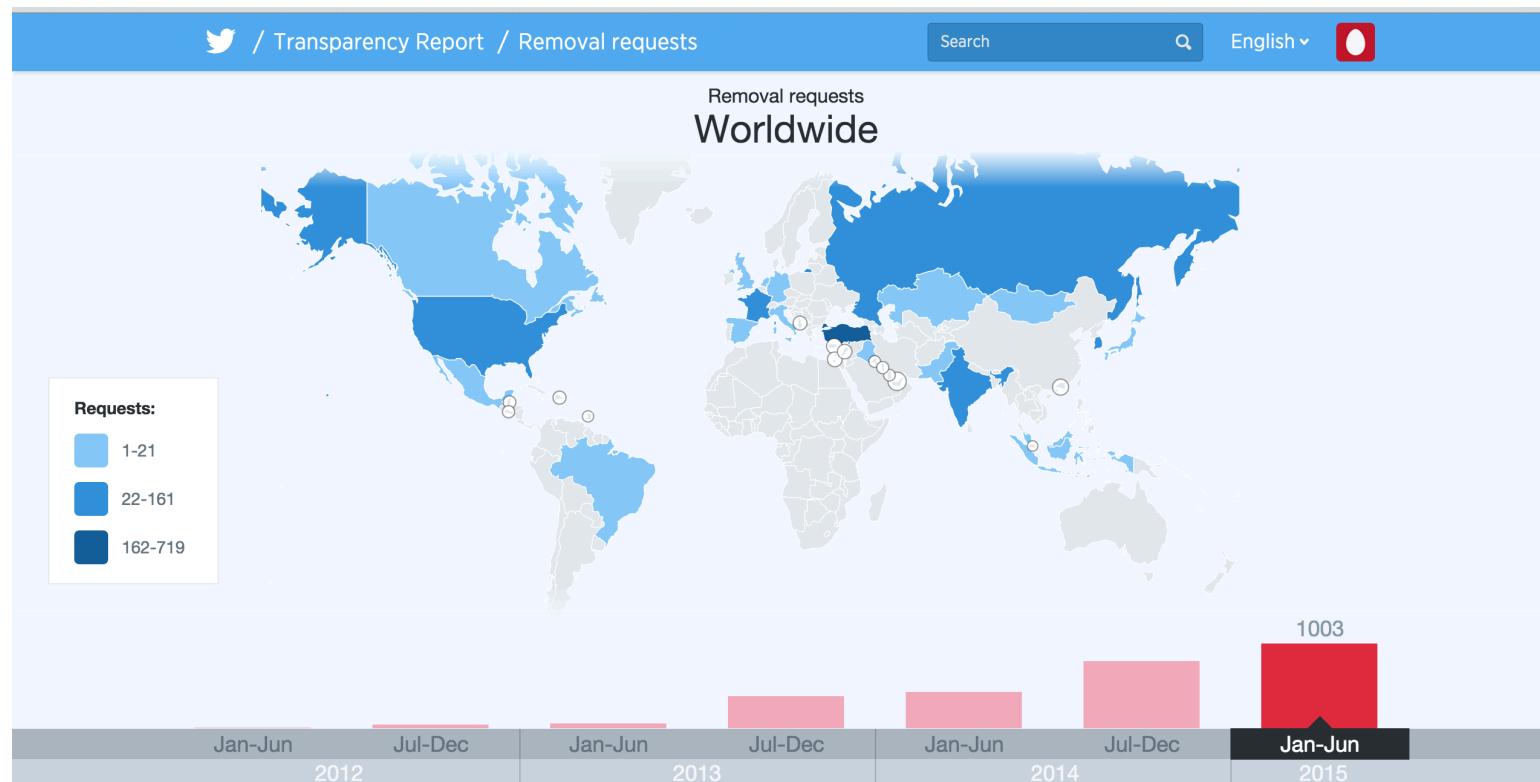
# Types of censorship

- **Network level:**
  - Block entire service.
  - China blocks western social media.
  - Turkey blocks Twitter March, 2014.
- **Application level:**
  - Service is not blocked, censorship is internal to the service
  - Example: *Sina Weibo* (Zhu et. al) – keyword filtering, etc.
  - **But what about application censorship in western social media?**

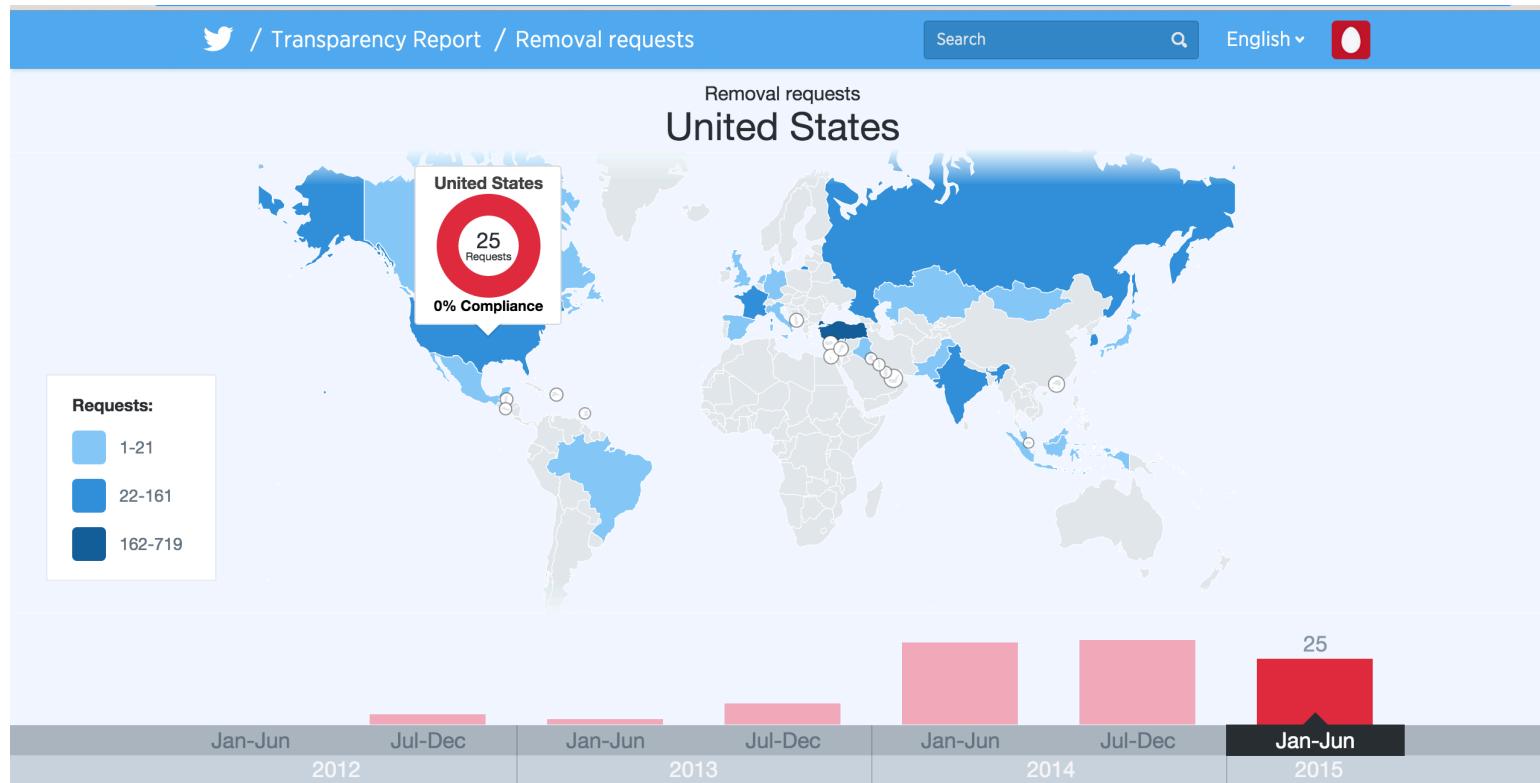
# Twitter censorship

- January 2012, Twitter “Country-Withheld Content”
- Twitter publishes censorship requests on ChillingEffects.org
- Twitter publishes bi-annual transparency reports

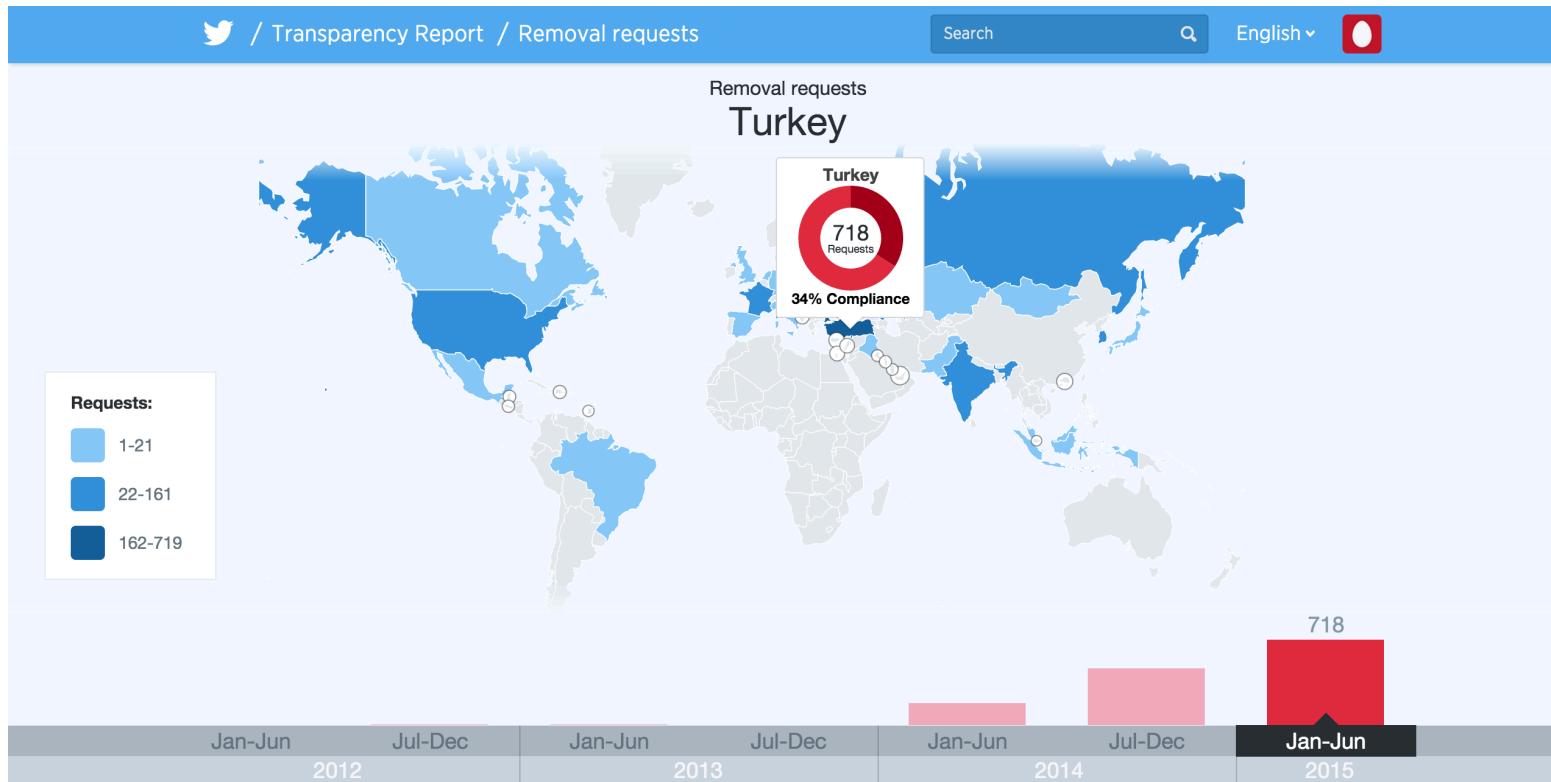
# Removal requests



# Removal requests USA



# Removal requests Turkey



# This map raises many questions

- What's up with Turkey?
- Why are there 0 requests from Arabic-speaking countries such as Saudi and UAE?
  - They have reputation for censorship.
  - Twitter is very popular in these countries.

# Are Twitter transparency reports complete?

- “NOTE: The data in these reports is as accurate as possible, but may not be 100% comprehensive.”
- Twitter does not post the notices to Chilling Effects when they “... are legally prohibited from doing so”

***Prohibited by whom? And which laws?***

# Request sample from Chilling Effects

T.C.  
ANKARA  
1. SULH CEZA HÂKİMLİĞİ

DEĞİŞİK İŞ KARAR

DEĞİŞİK İŞ NO : 2014/2759 D.İş

HAKİM : [REDACTED]  
KATİP : [REDACTED]

TALEP EDEN : [REDACTED]  
VEKİLİ : [REDACTED]  
KARŞI TARAFLI : www.twitter.com  
TALEP TARİHİ : 13/08/2014

Talep eden vekilinin tevzi sonucu hakimliğimize gönderilen dilekçesi okundu.

**GEREÇİ DÜŞÜNÜLDÜ:**

Talep eden [REDACTED] vekili [REDACTED]; 13/08/2014 tarihli  
dilekçesi ile

<https://www.twitter.com> adlı internet sitesinde  
<https://twitter.com/EmelTurangil/status/497773665489281024>

sayılı URL adresinde  
müvekkilinin kişilik haklarının ihlal edildiğinden bahisle erişimin engellenmesi  
talebinde bulunmuştur.

(Redactions by Twitter.)

# Why Turkey?

- It represents the darkest blob on the map.



- Turkey banned Twitter, in its entirely, in 2014.

# Research Questions

- Can we confirm the number of withheld tweets reported for Turkey in the Transparency Reports?
- Can we find unreported tweets in Turkey?
- Can we extract and analyze topics being withheld?

# Agenda

- Methodology
  - Validating censored tweets
  - Finding interesting users
  - Crawling
- Findings
- Topic analysis
- Bypassing censorship
- Future work

# How to validate if tweets are censored in Turkey?



# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.



# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy



# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy 



# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy 
  - Use the PlanetLab network



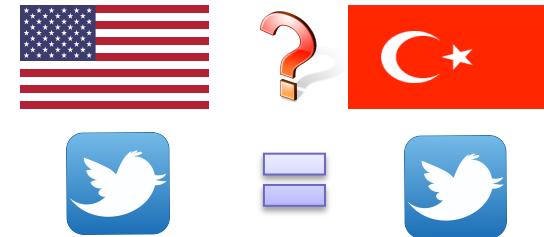
# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy 
  - Use the PlanetLab network 



# How to validate if tweets are censored in Turkey?

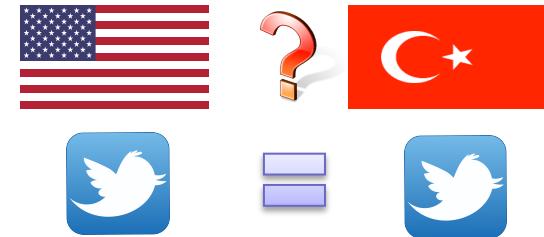
- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy 
  - Use the PlanetLab network 
- While analyzing a sample of known censored tweets, we observed a special field in the tweet structure, called:



```
"id_str" : "51707756700291...",  
"in_reply_to_user_id":null,  
"favorited":false,  
"withheld_in_countries": ["TR"] "...
```

# How to validate if tweets are censored in Turkey?

- View tweets from inside Turkey and check if they are invisible.
  - Use a free Turkish proxy 
  - Use the PlanetLab network 
- While analyzing a sample of known censored tweets, we observed a special field in the tweet structure, called:



```
"id_str" : "51707756700291...",  
"in_reply_to_user_id":null,  
"favorited":false,  
"withheld_in_countries": [ "TR" ] "...
```

- If this field is reliable, then we can crawl from home!

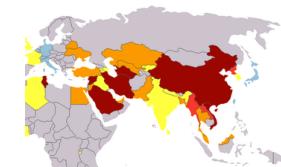
# Validating our observation

- Tor browser bundle with ExitNode {TR}.
- **Result:**
  - We confirmed that all tweets that contained the "withheld\_in\_countries" field, were indeed invisible in Turkey.
  - Therefore, can crawl from USA. ✓



# Crawling

- **Goal:** collect censored tweets and examine their content.
- Free Twitter public API.
- Start with 689 seed sensitive users:
  - Source: Chilling Effects website.
  - Spider-out in the social group for new users.
- API with geo bounding boxes of three major Turkish cities.
- Crawl historical users' timeline.



# Revisiting collected tweets

- Methodology: Collect now, verify later
  - Censorship doesn't happen immediately
  - Inspect if “withheld\_in\_countries” field is present
- Tradeoff: # users vs. time granularity

# Agenda

- Methodology
  - Validating censored tweets
  - Finding interesting users
  - Crawling
- **Results**
- Topic analysis
- Bypassing censorship
- Future work

# Results

Overall we collected	20+ mil tweets
Experiment duration (all methods)	10/2014-5/2015
Number of interesting users	7,642 users

# Results

Overall we collected	20+ mil tweets
Experiment duration (all methods)	10/2014-5/2015
Number of interesting users	7,642 users
Number of reported withheld tweets by 6/15	3,981 tweets
Number of reported withheld users by 6/15	204 users

# Results

Overall we collected	20+ mil tweets
Experiment duration (all methods)	10/2014-5/2015
Number of interesting users	7,642 users
Number of reported withheld tweets by 6/15	3,981 tweets
Number of reported withheld users by 6/15	204 users
<b>Number of withheld tweets we collected</b>	<b>266,407 tweets</b>
<b>Number of withheld users we identified</b>	<b>46 users</b>
<b>Number of withheld tweets not including those from withheld accounts</b>	<b>205,451 tweets</b>

# Results

**There is at least two orders of magnitude more withheld tweets in Turkey than what Twitter reported**

Number of reported withheld tweets by 6/15	3,981 tweets
Number of reported withheld accounts by 6/15	204 users
<b>Number of withheld tweets we collected</b>	<b>266,407 tweets</b>
<b>Number of withheld accounts we identified</b>	<b>46 users</b>
<b>Number of withheld tweets not including those from withheld accounts</b>	<b>205,451 tweets</b>

# Two orders of magnitude

- “NOTE: The data in these reports is as accurate as possible, but may not be 100% comprehensive.”
- **It is indeed not 100% comprehensive nor accurate.**

# Deduplication

- Maybe Twitter reports censored copies as one event?
- Note: the tweets we collected are unique by ID.
- We removed duplications: copy/paste & retweets (details in the paper).

# Deduplication

- Maybe Twitter reports censored copies as one event?
- Note: the tweets we collected are unique by ID.
- We removed duplications: copy/paste & retweets (details in the paper).
- Reduced the number of tweets to: **88,276**.

# Deduplication

- Maybe Twitter reports censored copies as one event?
- Note: the tweets we collected are unique by ID.
- We removed duplications: copy/paste & retweets (details in the paper).
- Reduced the number of tweets to: **88,276**.
- → **One order of magnitude higher than what Twitter reported.**

# Agenda

- Methodology
  - Validating censored tweets
  - Finding interesting users
  - Crawling
- Findings
- **Topic analysis**
- Bypassing censorship
- Future work

# Topic analysis

- Topic analysis is valuable for understanding the political aims of the Turkish censors.
- *Term frequency–inverse document frequency* (tf-idf) – standard machine learning algorithm.
- We extracted the top 5 topics with 10 words for each topic.

# Topic analysis

- Word: **media, dishonest, freedom, anti-Semitic references, minister, etc.**
- We can infer that strongly worded and vulgar political discussions are being targeted by Turkey's censorship authorities.

Turkish "topic"	English translation
doğan aydın medya ,serefsiz grubu medyası vatan hurriyet değil köpeği	Aydin Dogan (a person who owns the biggest media in Turkey) <b>media dishonest</b> group home <b>freedom</b> not <b>dog</b>
co rt ,sekerbank ,sap,sik in hikayesi dikkat chp ibrahim karaca	Not meaningful: Sekerbank (a Turkish financial institution) attention in story Ibrahim Karaca (a Turkish name)
koç un vehbi oğlu in aydın rahmi doğan yahudi nahum	<b>Vehbi Koç</b> (a Turkish entrepreneur / philanthropist) son enlightened womb rising <b>Jewish</b>
elvan lütfi bakan bakanı ula,stırma ttt aptal bi www adam	<b>Lütfi Elvan</b> (a Turkish government minister) <b>stupid man</b> <b>minister transportation</b>
davutoğlu ahmet ba,şbakan lan pic sikeyim yahudi gavat götür vatan	<b>Ahmet Davutoglu</b> (current Turkish prime minister) <b>prime minister</b> man <b>bastard fuck Jewish pimp ass</b>

# Agenda

- Methodology
  - Validating censored tweets
  - Finding interesting users
  - Crawling
- Findings
- Topic analysis
- **Bypassing censorship**
- Future work

# Bypassing censorship

- In April 2015, we followed a group of withheld accounts in Turkey, and noticed that some users were still tweeting from inside the country despite their accounts being withheld.

# Bypassing censorship

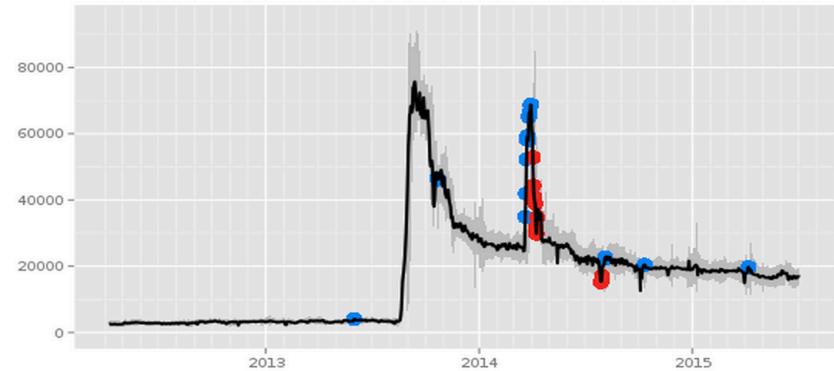
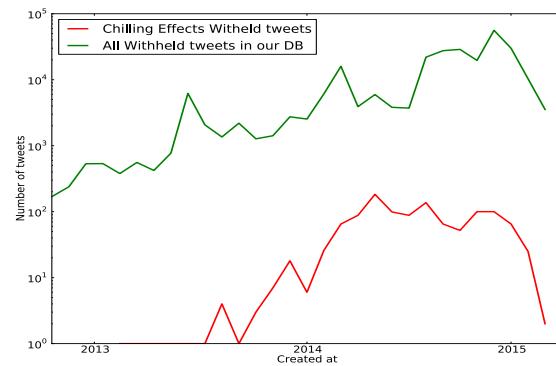
- In April 2015, we followed a group of withheld accounts in Turkey, and noticed that some users were still tweeting from inside the country despite their accounts being withheld.
- **How?**

# Bypassing censorship

- In April 2015, we followed a group of withheld accounts in Turkey, and noticed that some users were still tweeting from inside the country despite their accounts being withheld.
- **How?**
- VPN & Tor.

# Users connecting to Tor in Turkey

1. Tor metric: Number of users connecting to Tor in Turkey daily.
2. Distribution of withheld tweet by month (log scale).

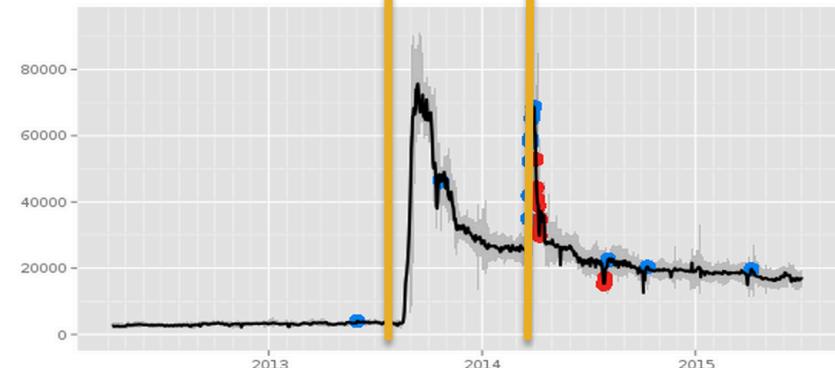
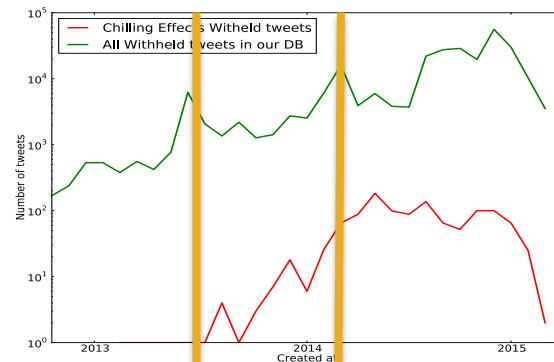


# Users connecting to Tor in Turkey

1. Tor metric: Number of users connecting to Tor in Turkey daily.
2. Distribution of withheld tweet by month (log scale).

## Peaks:

1. 5/2013 Taksim square protest.
2. 3/2014 Twitter ban in Turkey.

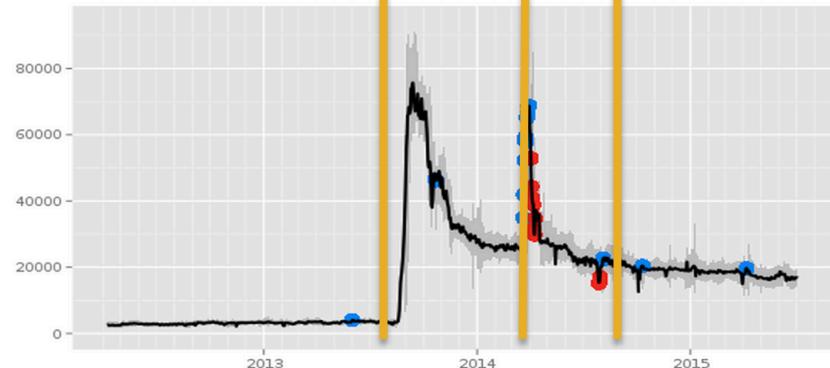
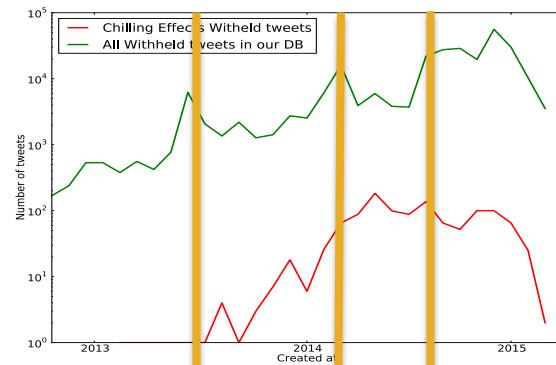


# Users connecting to Tor in Turkey

1. Tor metric: Number of users connecting to Tor in Turkey daily.
2. Distribution of withheld tweet by month (log scale).

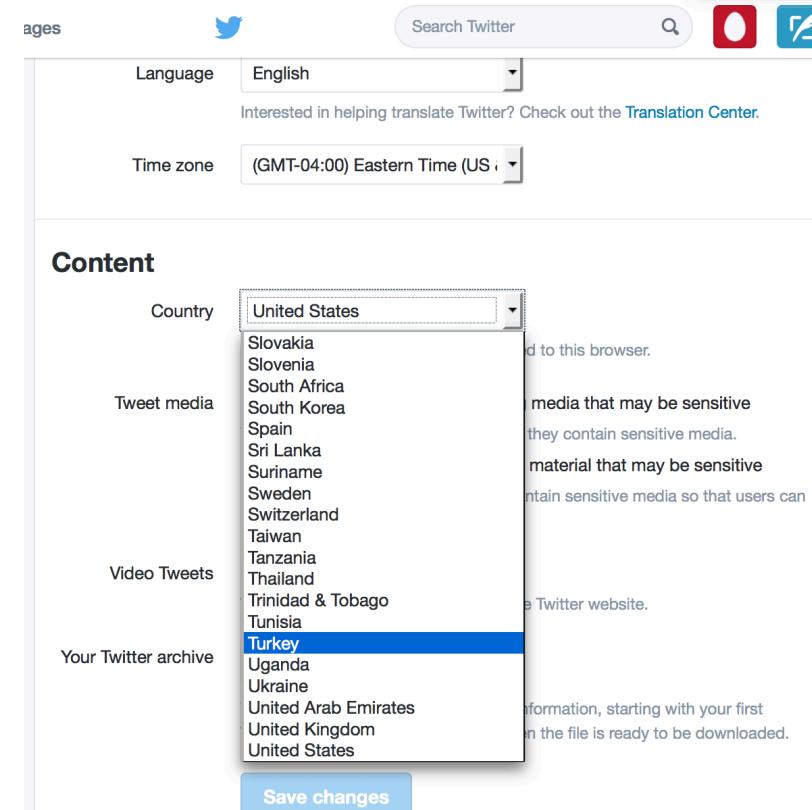
## Peaks:

1. 5/2013 Taksim square protest.
2. 3/2014 Twitter ban.
3. 11/2014 &12/2014: peak in “withheld content” censorship



# Bypassing censorship

- Turns out, it is even easier!
- Change the “Location” setting in the Twitter application to “USA” for example.
- We expect many users to start using this method in the future.



# What if Twitter becomes vigilant?

- Users will revert to VPNs or Tor.

# Country level censorship is hard

- If “country withholding” mechanisms don’t work, countries will demand global Twitter censorship

French data regulator rejects Google's right-to-be-forgotten appeal

Search engine's attempt to block French order to apply delistings to its google.com domain, not just its European sites, dismissed in ground-breaking case



# Future work

- Measuring censorship in other countries.
  - Twitter imposes restrictive rate limits.
- Political science collaboration ongoing.

# Conclusion

1. We provided methods to find unpublished censored tweets
2. We showed that the size of censorship in Turkey is at least two orders of magnitude larger than what Twitter reported.
3. We introduced a new simple method to bypass Twitter censorship by changing the location setting.
4. We extracted censored topics using machine learning clustering algorithm and found that most of censored topics are political.

FAQ's:

<http://www.cs.rice.edu/~rst5/twitterTurkey/>