Max
Planck
Institute
for
Software Systems

Forgetting in Social Media: Understanding and Controlling Longitudinal Exposure of Socially Shared Data

Mainack Mondal[†] Johnnatan Messias[†] Saptarshi Ghosh[*]

Krishna P. Gummadi[†] Aniket Kate[‡]

[†]MPI-SWS [*]IIEST Shibpur [‡]Purdue University

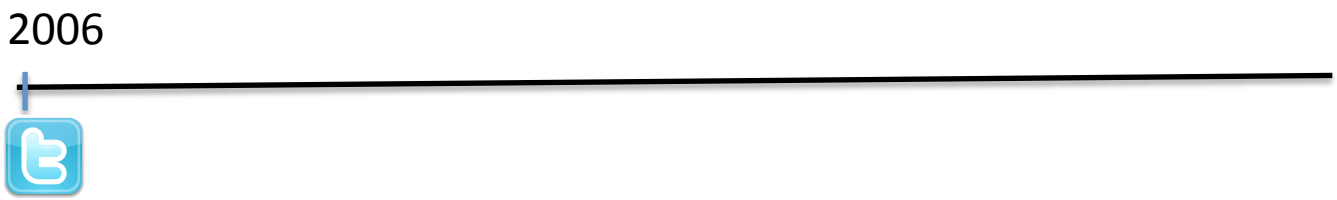
SOUPS 2016

Online Social Media sites (OSMs) are aging

OSMs are already around for a decade

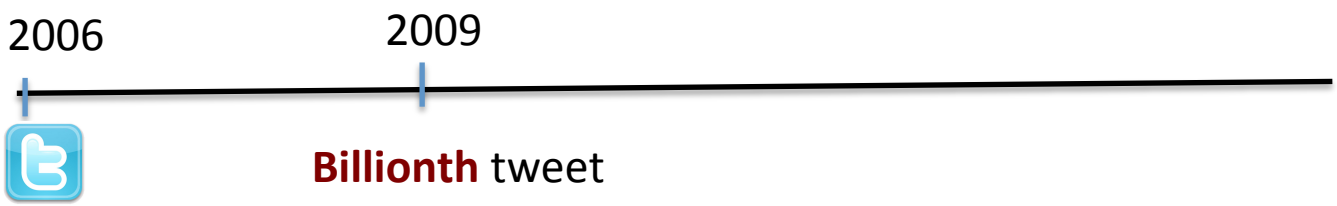
Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



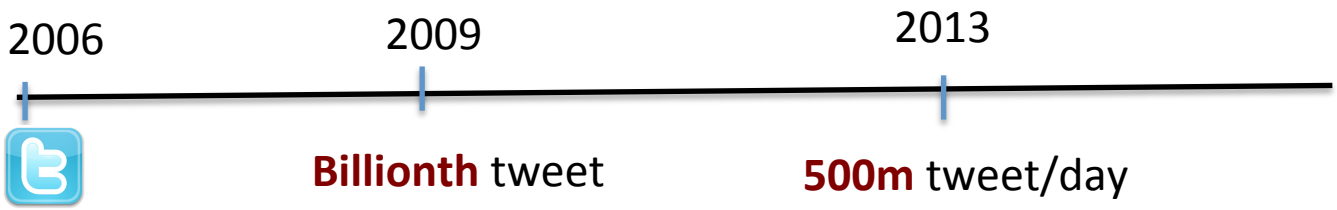
Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



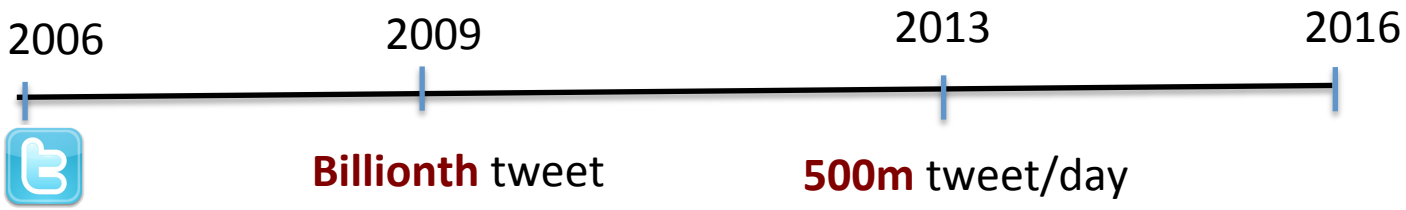
Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



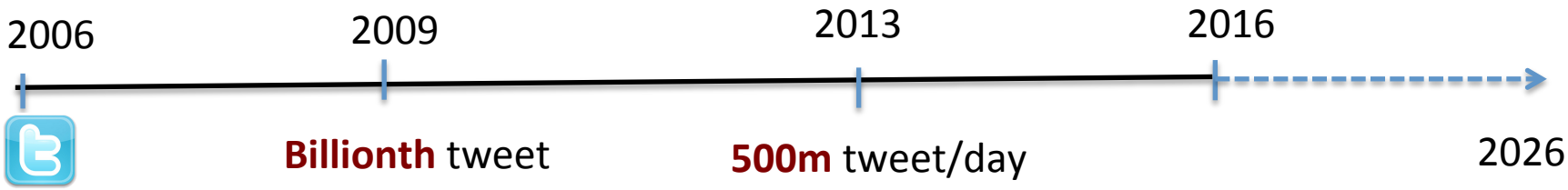
Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



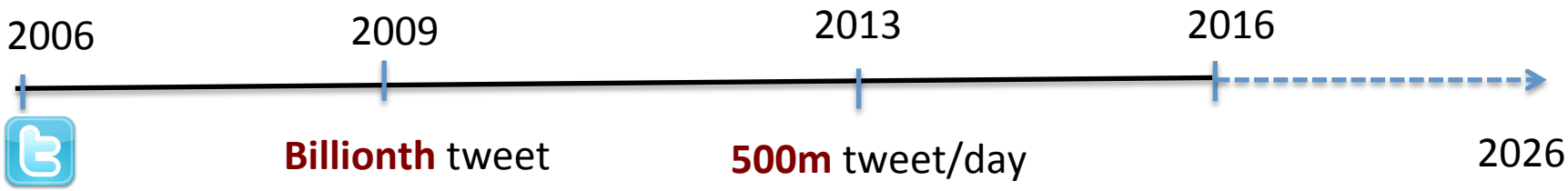
Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



Online Social Media sites (OSMs) are aging

OSMs are already around for a decade



In sites like Twitter

- Users are **content creators and managers**
- They might even need **to change privacy preferences over time**

Users change privacy preferences over time

2009

Content posted in freshman year: shared with everybody on internet

2012

3 years later: Hiring manager and colleagues **should not** see this

Users change privacy preferences over time

2009

Content posted in freshman year: shared with everybody on internet

2012

3 years later: Hiring manager and colleagues **should not** see this

They need to control **longitudinal exposure**: control **who can see old content**

Understanding longitudinal exposure control

[WPES 2013]
[SOUPS 2013]

Recent studies found via user surveys

- Users' willingness to share content drops as the content become old
- Willingness of share further decreases with a life-change

Understanding longitudinal exposure control

Recent studies found via user surveys

- Users' willingness to share content drops as the content become old
- Willingness of share further decreases with a life-change

A large scale study on tweets posted within a week reported

- 2.4% of those tweets are deleted by users within their week of observation
- However **they only considered** content posted in **very recent past**

[CSCW 2013]

Understanding longitudinal exposure control

Recent studies found via user surveys

- Users' willingness to share content drops as the content become old
- Willingness of share further decreases with a life-change

A large scale study on tweets posted within a week reported

- 2.4% of those tweets are deleted by users within their week of observation
- However **they only considered** content posted in **very recent past**

No investigation so far about

- Do users **change privacy preferences** to control **longitudinal exposure?**
- How **effective** are **current mechanisms** to control longitudinal exposure?

Goal

To better understand and control longitudinal exposure in OSMs

Rest of the talk

- ✔ Do users change privacy preferences over time?
- ✔ How effective are these exposure control mechanisms?
- ✔ How can these mechanisms be improved?

Rest of the talk

✔ **Do users change privacy preferences over time?**

✔ How effective are these exposure control mechanisms?

✔ How can these mechanisms be improved?

Collecting data on users changing privacy preferences

In this study we focus on Twitter 

Simple privacy preferences

- Either publicly visible to everyone
- Or withdrawn from public domain (by deletion or making account private)

Collecting data on users changing privacy preferences

In this study we focus on Twitter 

Simple privacy preferences

- Either publicly visible to everyone
- Or withdrawn from public domain (by deletion or making account private)

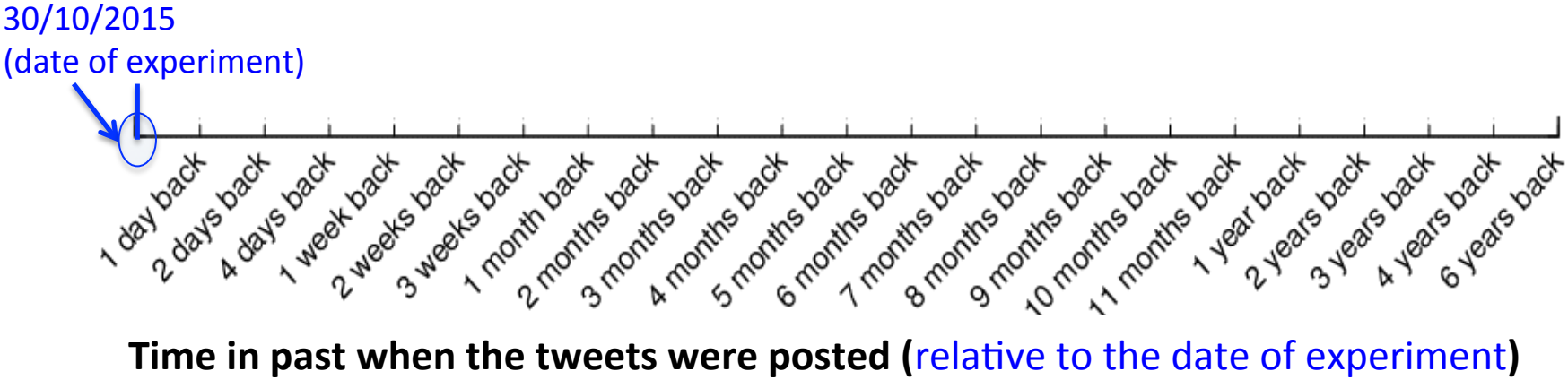
30/10/2015
(date of experiment)

Collecting data on users changing privacy preferences

In this study we focus on Twitter

Simple privacy preferences

- Either publicly visible to everyone
- Or withdrawn from public domain (by deletion or making account private)



Time in past when the tweets were posted (relative to the date of experiment)

Collecting data on users changing privacy preferences

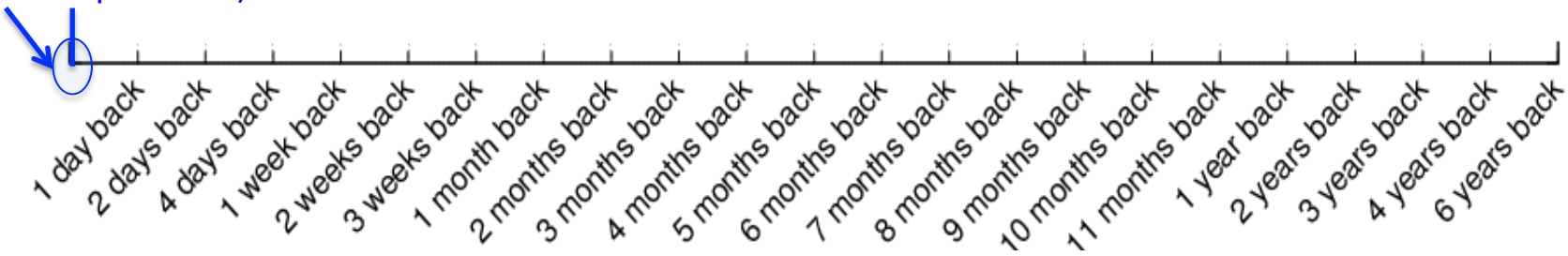
In this study we focus on Twitter

Simple privacy preferences

Either publicly visible to everyone

Or withdrawn from public domain (by deletion or making account private)

30/10/2015
(date of experiment)



Time in past when the tweets were posted (relative to the date of experiment**)**
ranges from tweets posted **1 day back** to **6 years back**

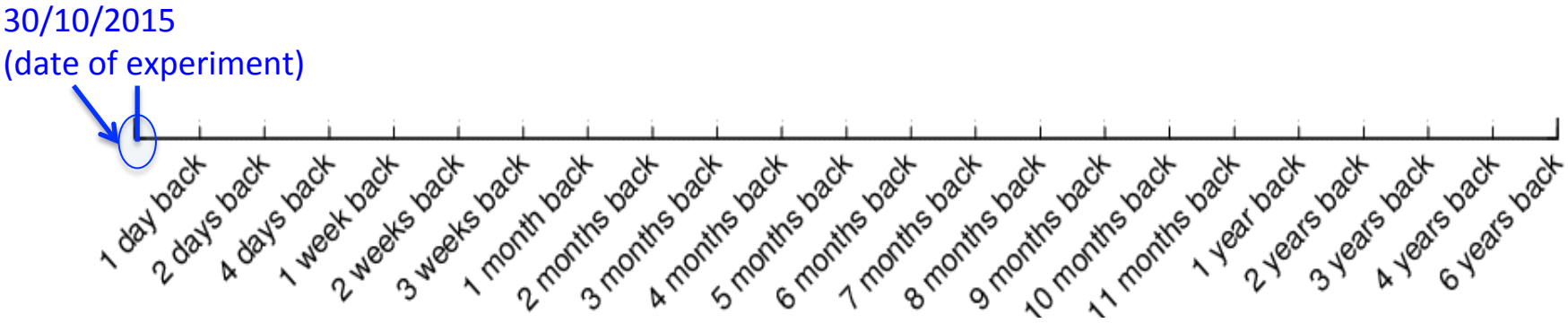
Collecting data on users changing privacy preferences

In this study we focus on Twitter

Simple privacy preferences

Either publicly visible to everyone

Or withdrawn from public domain (by deletion or making account private)



Time in past when the tweets were posted (relative to the date of experiment)
ranges from tweets posted **1 day back** to **6 years back**

All of these past tweets were **public when they were posted**

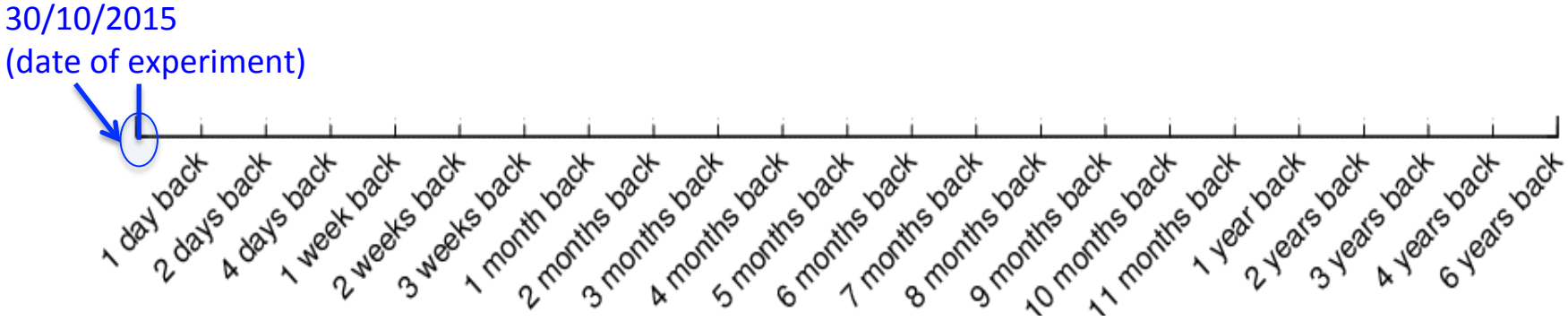
Collecting data on users changing privacy preferences

In this study we focus on Twitter

Simple privacy preferences

Either publicly visible to everyone

Or withdrawn from public domain (by deletion or making account private)

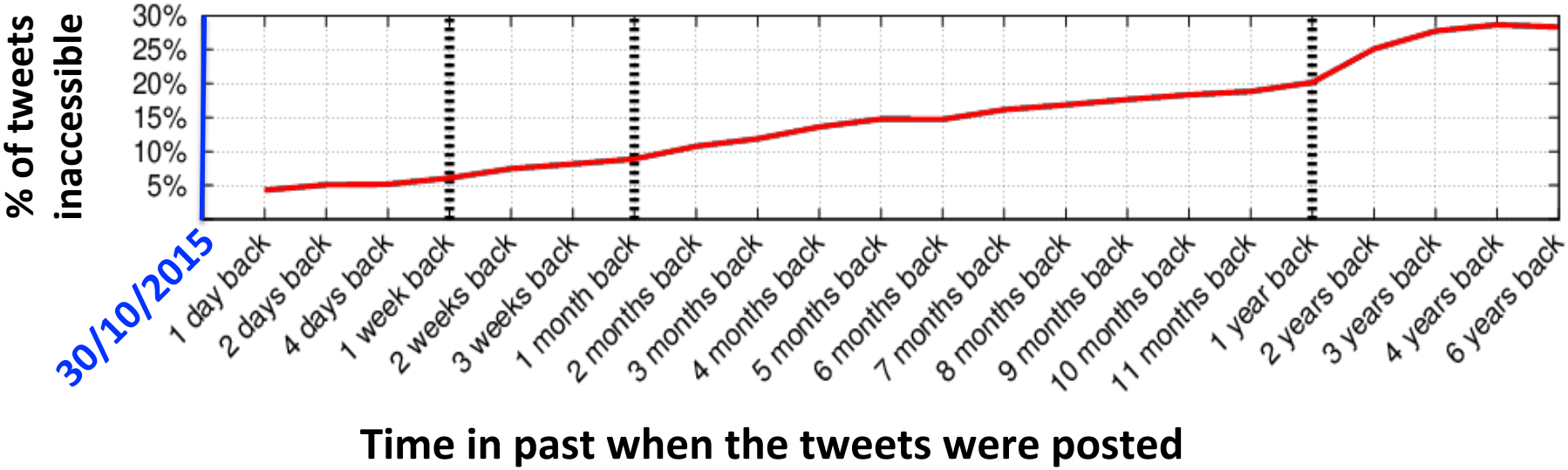


Time in past when the tweets were posted (relative to the date of experiment**)**
ranges from tweets posted **1 day back** to **6 years back**

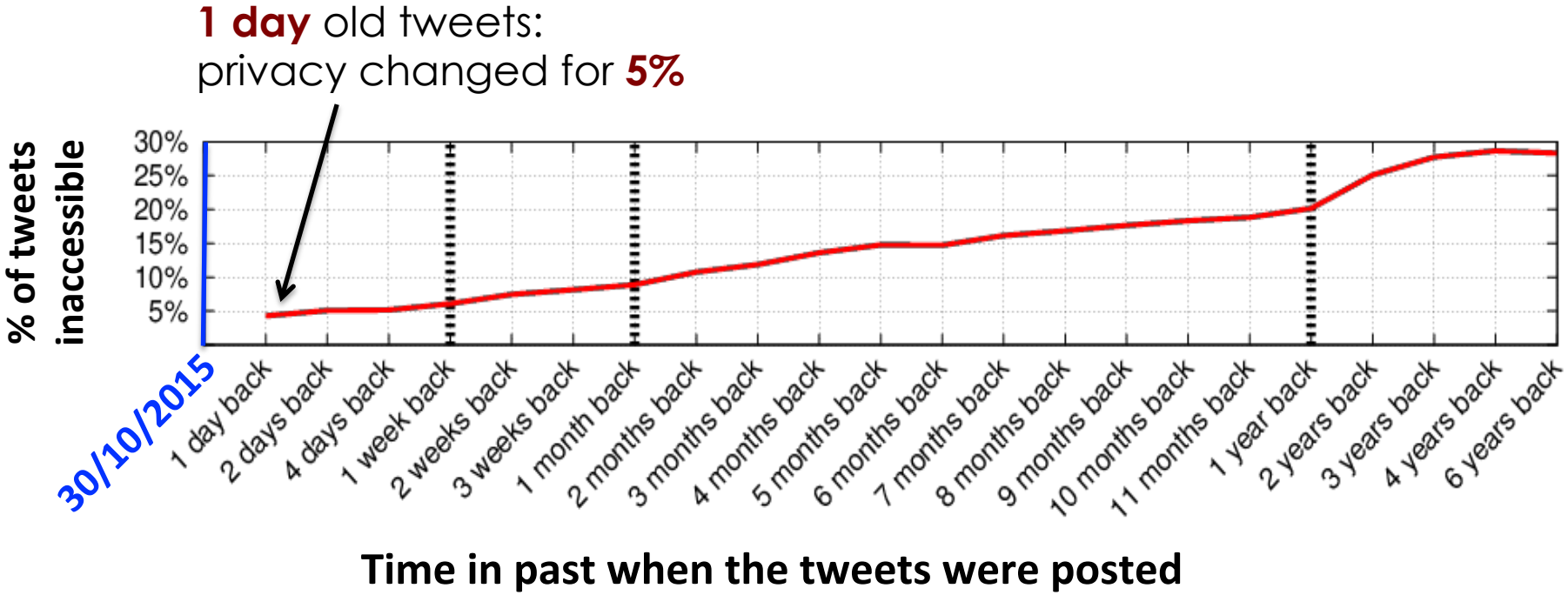
All of these past tweets were **public when they were posted**

If **inaccessible** on experiment date, privacy preferences **changed** over time

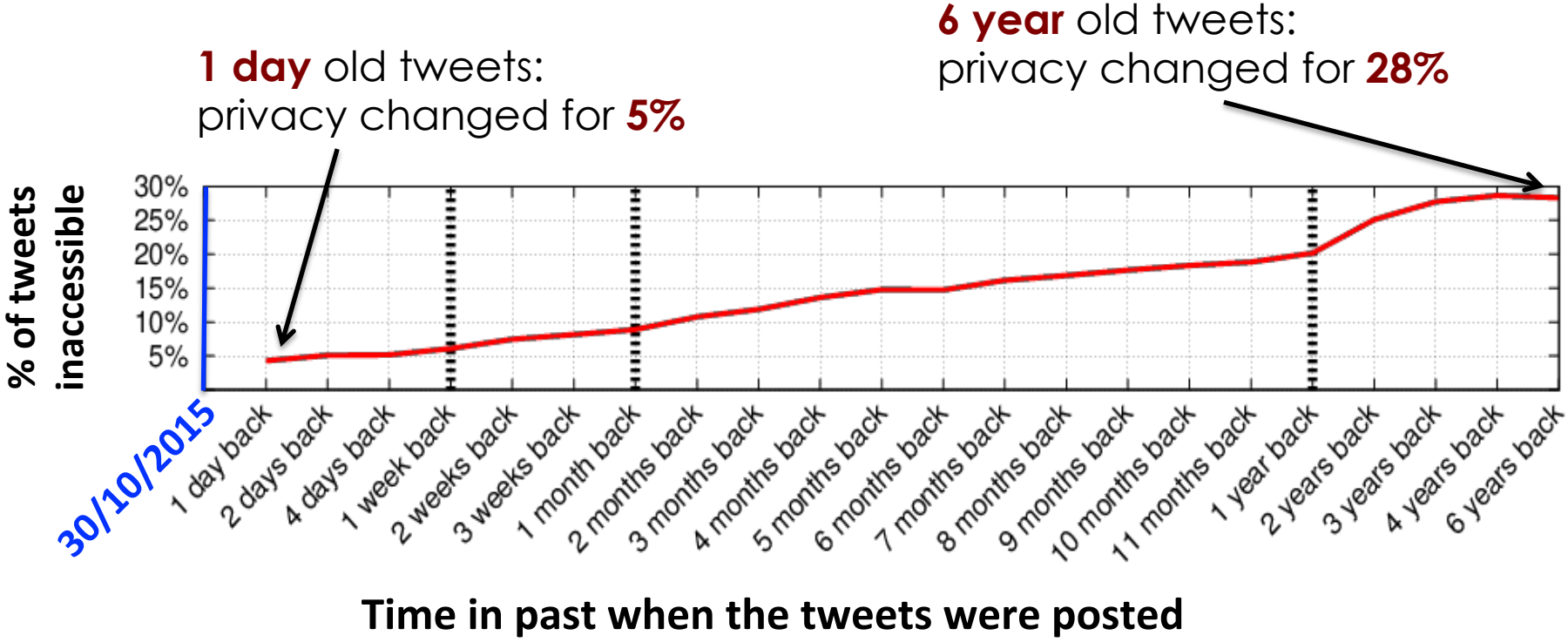
Do users change privacy preferences over time?



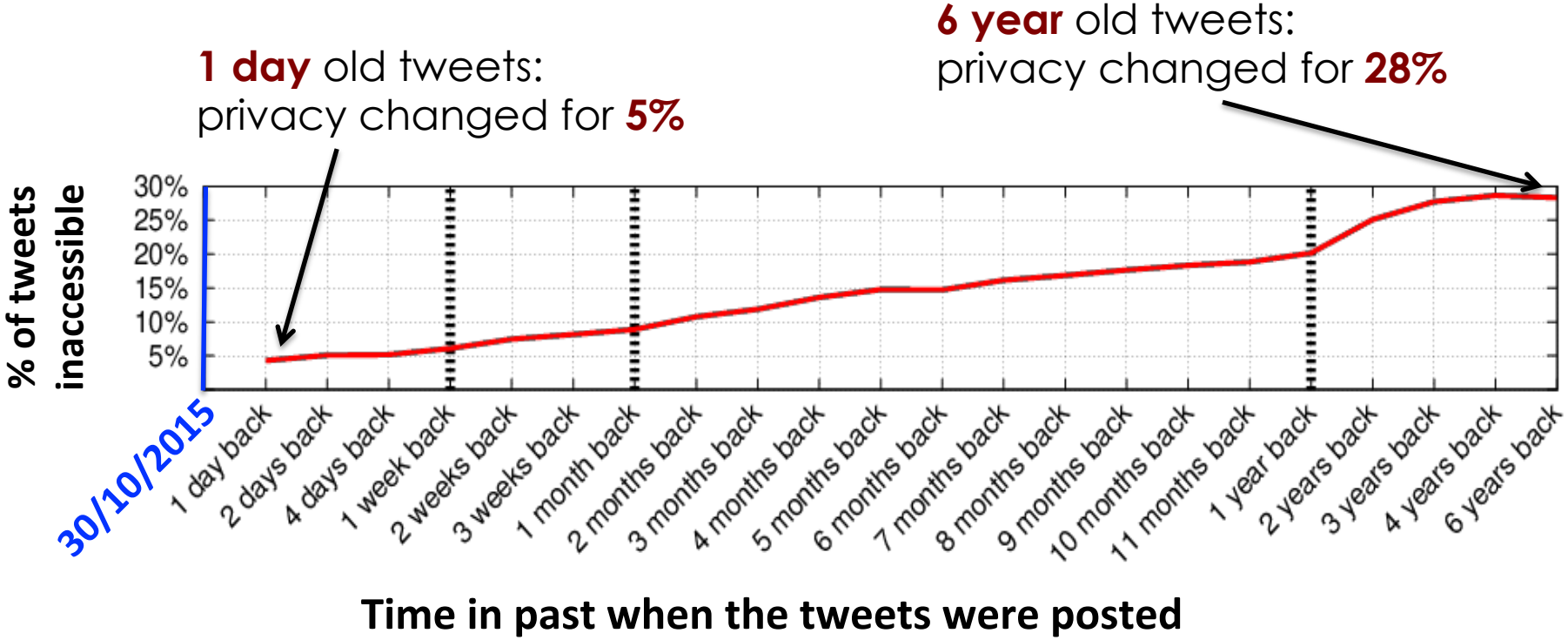
Do users change privacy preferences over time?



Do users change privacy preferences over time?

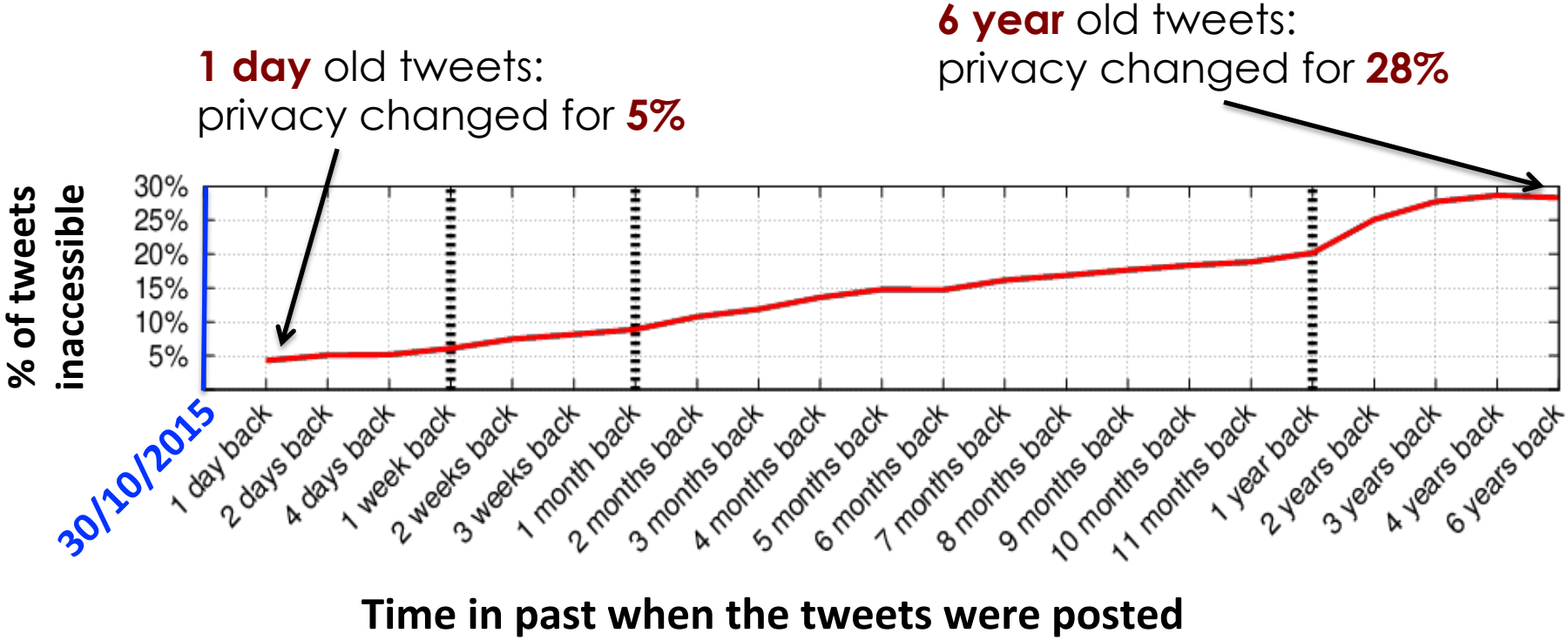


Do users change privacy preferences over time?



Users change privacy for increasing amount of old data with time

Do users change privacy preferences over time?



Users change privacy for increasing amount of old data with time

How do these users change privacy of this content?

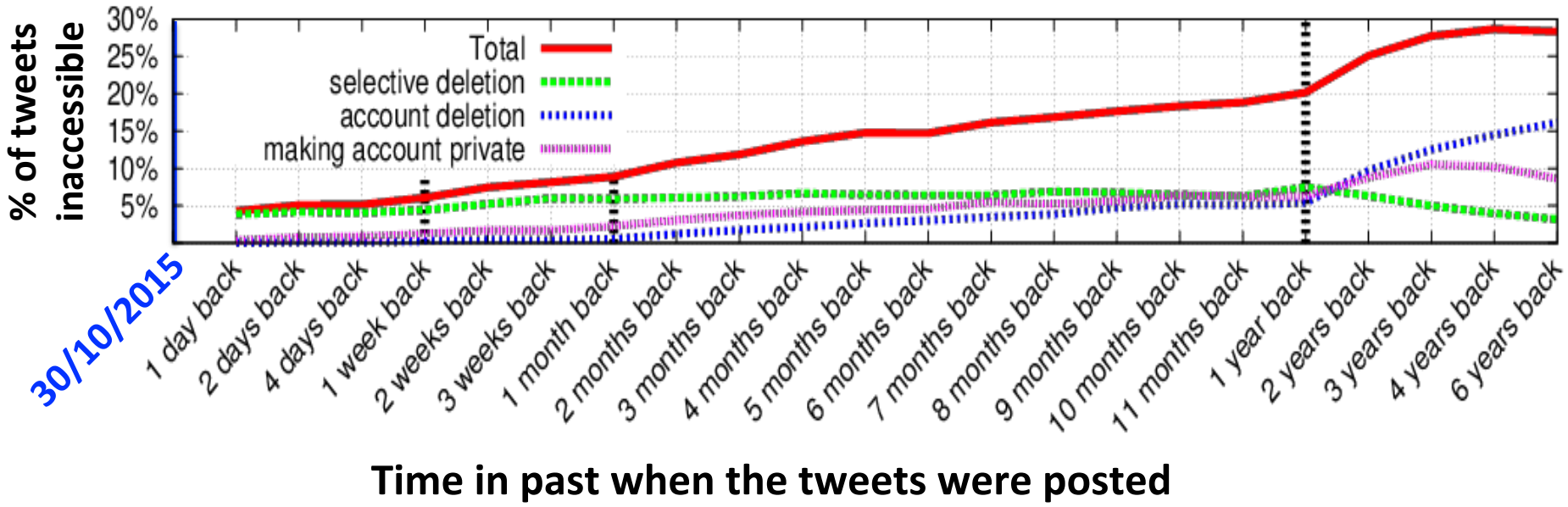
Mechanisms to change privacy preferences in Twitter

Three ways users change privacy of old content in Twitter

They are the longitudinal exposure control mechanisms

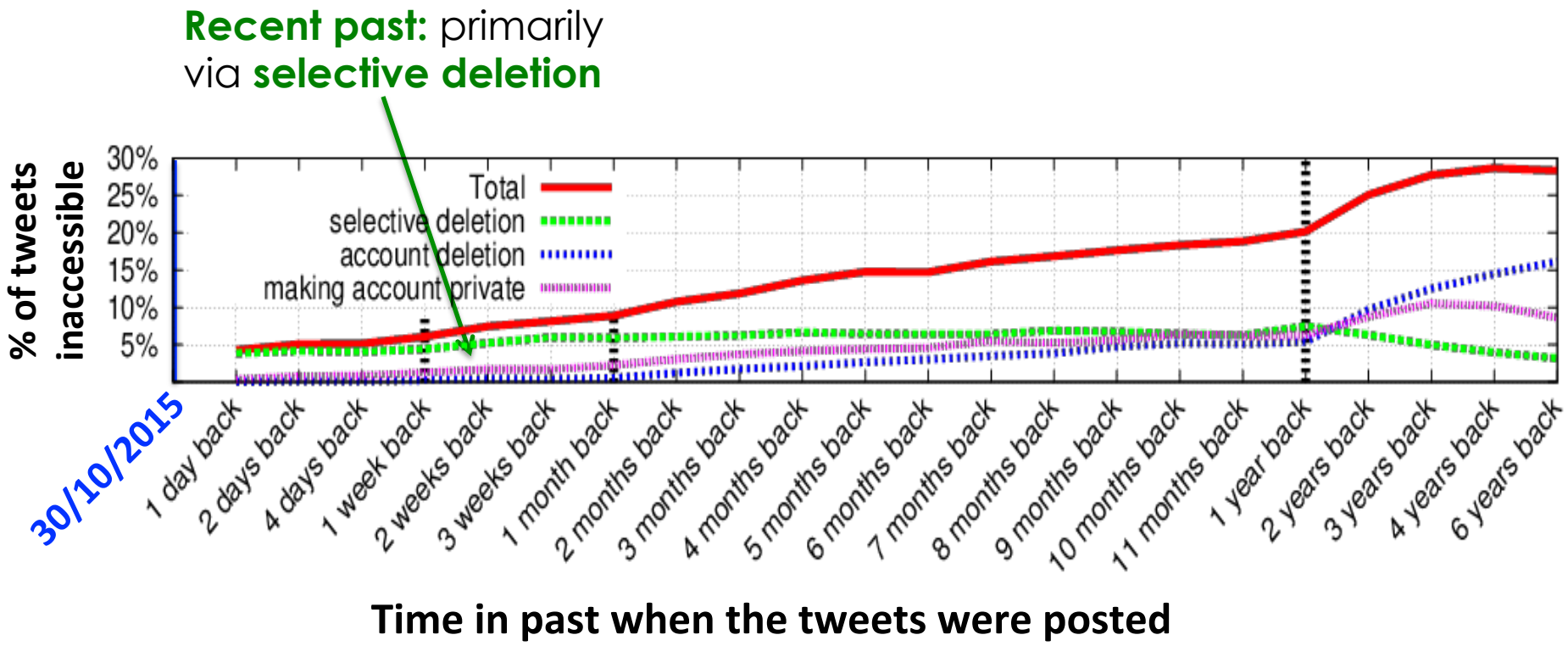
Mechanism	Description
Selective deletion	Selectively withdraw some old tweets to control exposure
Account deletion	Withdraw all old tweets to control exposure in bulk
Making account private	Withdraw all old tweets to control exposure in bulk

How do users change privacy preferences?

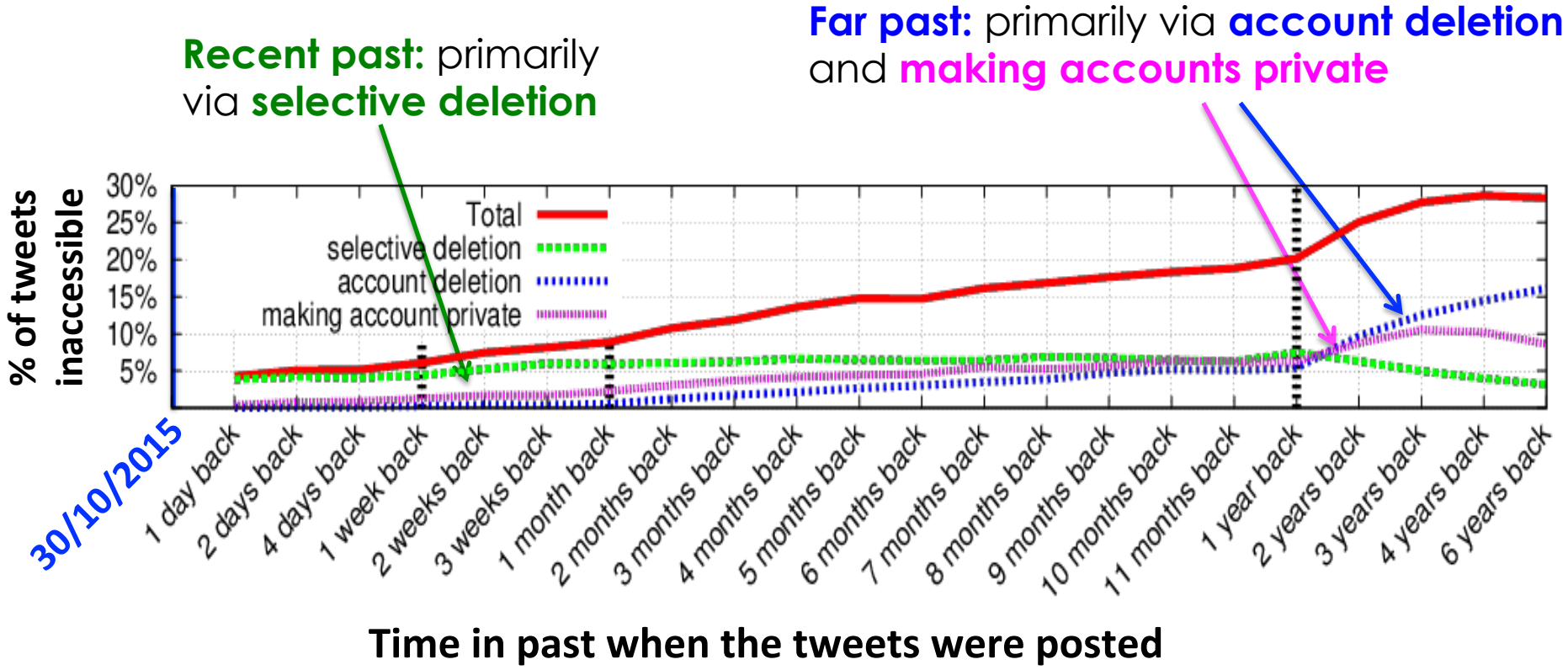


Time in past when the tweets were posted

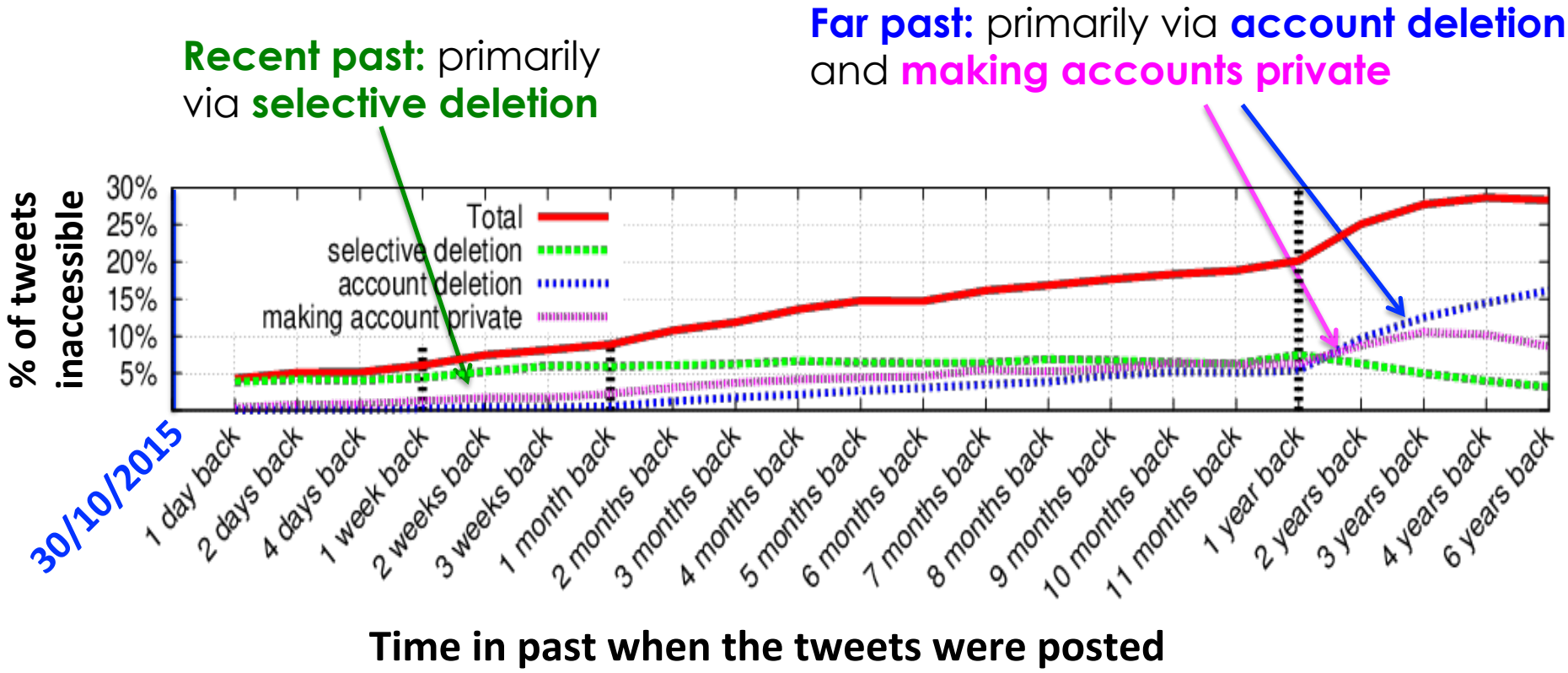
How do users change privacy preferences?



How do users change privacy preferences?



How do users change privacy preferences?



Very different mechanisms to change privacy for content from far past compared to recent past

Do many users change privacy of old content?

We randomly sample **100k** active users from 2009

Out of 8.9m random old tweets from these users 29.1% is inaccessible

What fraction of users change privacy of their content?

Do many users change privacy of old content?

We randomly sample **100k** active users from 2009

Out of 8.9m random old tweets from these users 29.1% is inaccessible

What fraction of users change privacy of their content?

User type	% of all users
Selectively deleted tweets	8.3%
Deleted their account	15.9%
Made their account private	10.4%
Users who took actions that changed privacy of their content	**34.6%**

Do many users change privacy of old content?

We randomly sample **100k** active users from 2009

Out of 8.9m random old tweets from these users 29.1% is inaccessible

What fraction of users change privacy of their content?

User type	% of all users
Selectively deleted tweets	8.3%
Deleted their account	15.9%
Made their account private	10.4%
Users who took actions that changed privacy of their content	**34.6%**

Do many users change privacy of old content?

We randomly sample **100k** active users from 2009

Out of 8.9m random old tweets from these users 29.1% is inaccessible

What fraction of users change privacy of their content?

User type	% of all users
Selectively deleted tweets	8.3%
Deleted their account	15.9%
Made their account private	10.4%
Users who took actions that changed privacy of their content	**34.6%**

A significant fraction of users change privacy of their old content

Demographics of users changing privacy

We investigated the demographics of our users from 2009

Inferred **Gender** and location form Twitter profile

Demographics of users changing privacy

We investigated the demographics of our users from 2009

Inferred **Gender** and location form Twitter profile

User type	% female users
Random sample of Twitter users	50.3%
Users who **did not delete** any content	**44.5%**
Deleted tweet selectively	**55.7%**
Deleted account/Made **account private**	**61.5%**

Demographics of users changing privacy

We investigated the demographics of our users from 2009

Inferred **Gender** and location form Twitter profile

User type	% female users
Random sample of Twitter users	50.3%
Users who **did not delete** any content	**44.5%**
Deleted tweet selectively	**55.7%**
Deleted account/Made **account private**	**61.5%**

Demographics of users changing privacy

We investigated the demographics of our users from 2009

Inferred **Gender** and location form Twitter profile

User type	% female users
Random sample of Twitter users	50.3%
Users who **did not delete** any content	**44.5%**
Deleted tweet selectively	**55.7%**
Deleted account/Made **account private**	**61.5%**

Demographics of users changing privacy

We investigated the demographics of our users from 2009

Inferred **Gender** and location form Twitter profile

User type	% female users
Random sample of Twitter users	50.3%
Users who **did not delete** any content	**44.5%**
Deleted tweet selectively	**55.7%**
Deleted account/Made **account private**	**61.5%**

Female users are more likely to **change privacy** of old content

Rest of the talk

✔ Do users change privacy preferences over time?

Privacy preferences are changed for significant fraction of old content

✔ **How effective are these exposure control mechanisms?**

✔ How can these mechanisms be improved?

Limitation of current exposure control: Residual activities

Limitation of current exposure control: Residual activities

Limitation of current exposure control: Residual activities

Limitation of current exposure control: Residual activities

These **conversations from** other users **remain public** even **after** a user **remove** her tweets/account

Limitation of current exposure control: Residual activities

These **conversations from** other users **remain public** even **after** a user **remove** her tweets/account

We call these **conversations residual activities**

Residual activities **contain information** about **withdrawn old content**

Anybody online can collect and analyze them by a username search

Limitation of current exposure control: Residual activities

These **conversations from** other users **remain public** even **after** a user **remove** her tweets/account

We call these **conversations residual activities**

Residual activities **contain information** about **withdrawn old content**

Anybody online can collect and analyze them by a username search

Residual activities might breach longitudinal exposure control

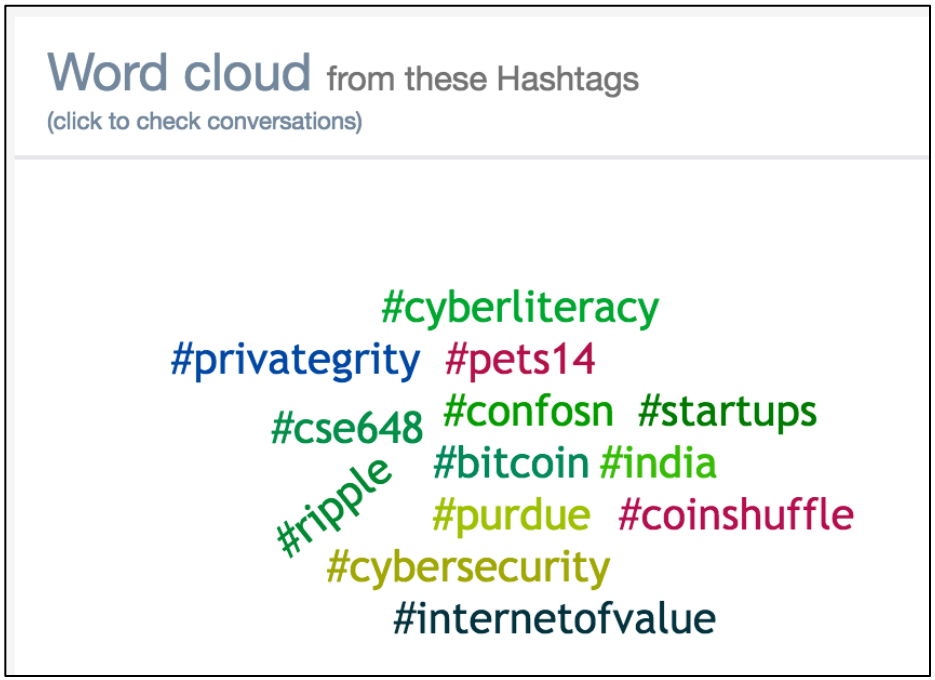
What information can we recover from residual activities?

User interests revealed by residual activities

Residual activities reveal **hashtags** which identify likely **user interests**

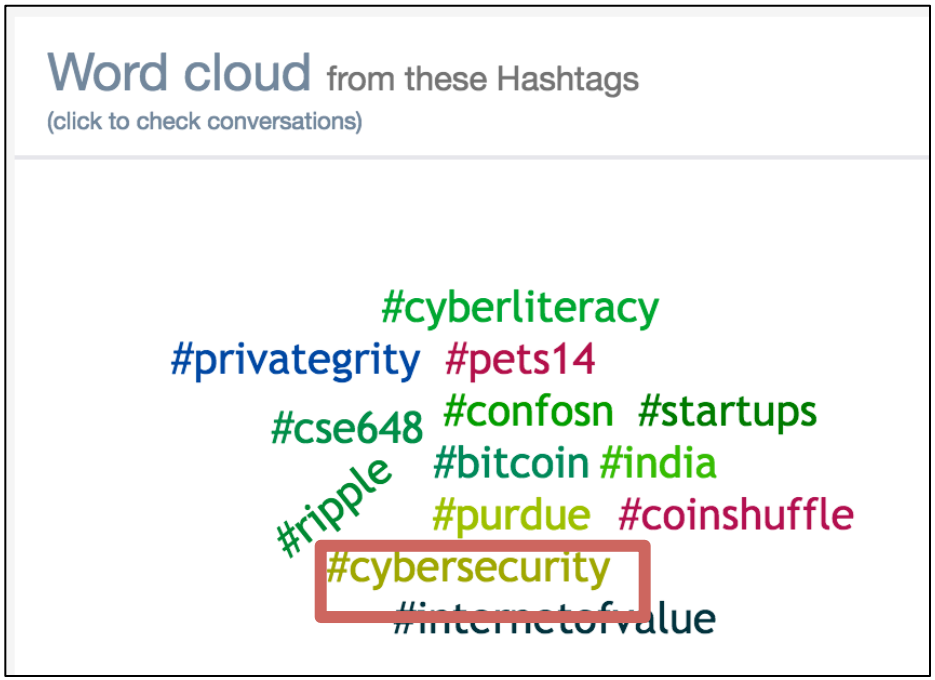
User interests revealed by residual activities

Residual activities reveal **hashtags** which identify likely **user interests**



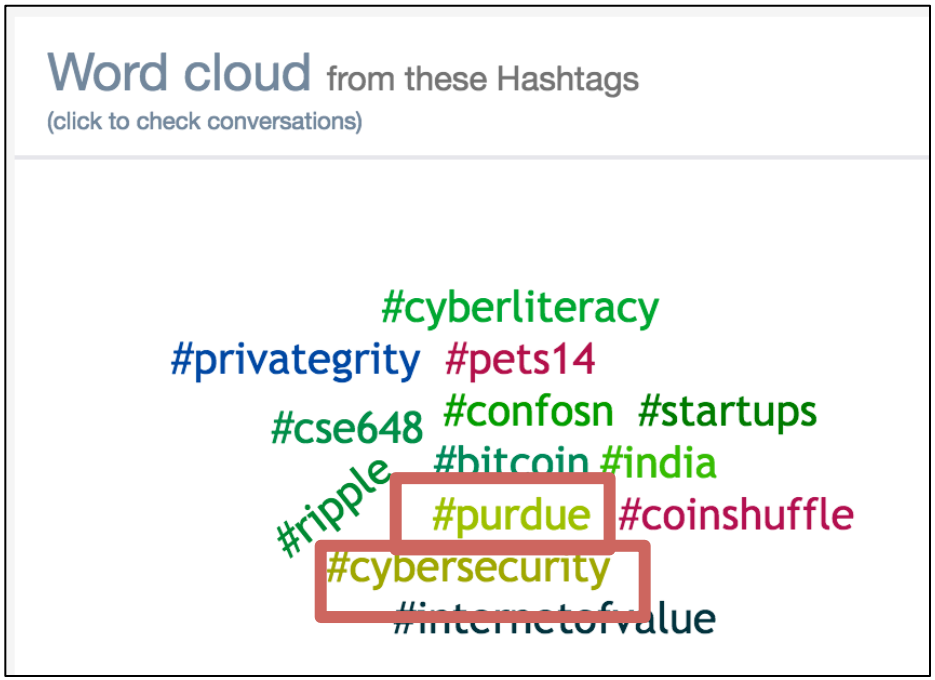
User interests revealed by residual activities

Residual activities reveal **hashtags** which identify likely **user interests**



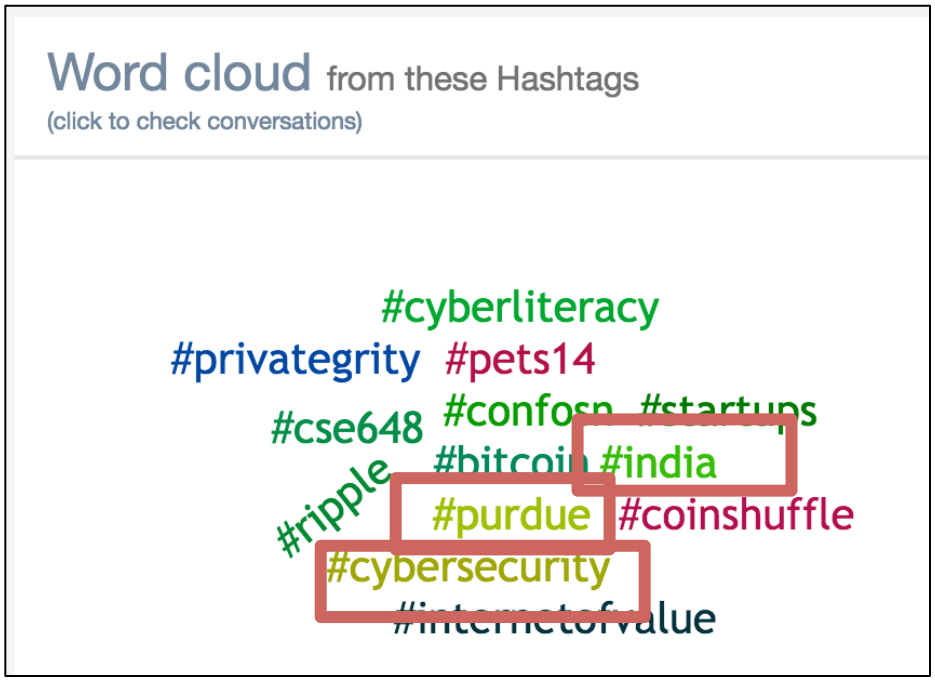
User interests revealed by residual activities

Residual activities reveal **hashtags** which identify likely **user interests**



User interests revealed by residual activities

Residual activities reveal **hashtags** which identify likely **user interests**



Sensitive user interests revealed by residual activities

We checked user interests revealed for deleted/private accounts from 2009

Sensitive user interests revealed by residual activities

We checked user interests revealed for deleted/private accounts from 2009

Deleted/private accounts	Topics of interest from hashtags	Hashtags revealed by residual activities
Account 1	Politics, Sports, Technology	#iranelection, #prisoners, #strike, #frenchopen, #tech
Account 2	Sports, LGBTQ issues	#daviscup, #samesexsunday, #india, #lgbt, #followfriday
Account 3	Sports	#grandrapids, #nascar

Sensitive user interests revealed by residual activities

We checked user interests revealed for deleted/private accounts from 2009

Deleted/private accounts	Topics of interest from hashtags	Hashtags revealed by residual activities
Account 1	Politics, Sports, Technology	#iranelection, #prisoners, #strike, #frenchopen, #tech
Account 2	Sports, LGBTQ issues	#daviscup, #samesexsunday #india, #lgbt, #followfriday
Account 3	Sports	#grandrapids, #nascar

Sensitive user interests revealed by residual activities

We checked user interests revealed for deleted/private accounts from 2009

Deleted/private accounts	Topics of interest from hashtags	Hashtags revealed by residual activities
Account 1	Politics, Sports, Technology	#iranelection, #prisoners, #strike, #frenchopen, #tech
Account 2	Sports, LGBTQ issues	#daviscup, #samesexsunday #india, #lgbt, #followfriday
Account 3	Sports	#grandrapids, #nascar

Some of these hashtags can be considered **sensitive**

Sensitive user interests revealed by residual activities

We checked user interests revealed for deleted/private accounts from 2009

Deleted/private accounts	Topics of interest from hashtags	Hashtags revealed by residual activities
Account 1	Politics, Sports, Technology	#iranelection, #prisoners, #strike, #frenchopen, #tech
Account 2	Sports, LGBTQ issues	#daviscup, #samesexsunday #india, #lgbt, #followfriday
Account 3	Sports	#grandrapids, #nascar

Some of these hashtags can be considered **sensitive**

Residual activities also reveal

- Demographics of accounts
- Meaning of deleted tweets -- Check out our paper for details

Residual activities can leak information about withdrawn accounts/tweets and breach longitudinal exposure control

Residual activities can leak information about withdrawn accounts/tweets and breach longitudinal exposure control

We developed a web app for users to check residual activities

Check out the app is at: **http://twitter-app.mpi-sws.org/footprint/**

Rest of the talk

✔ Do users change privacy preferences over time?

Privacy preferences are changed for significant fraction of old content

✔ How effective are these exposure control mechanisms?

Current mechanisms do not take care of information leakage by residual activities

✔ **How can these mechanisms be improved?**

Dealing with residual activities is difficult

Straw man:

Withdraw all the residual activities with original tweet/account

Problem

Residual activities are **not "owned"** by the original poster

Dealing with residual activities is difficult

Straw man:

Withdraw all the residual activities with original tweet/account

Problem

Residual activities are **not "owned"** by the original poster

Emerging OSMs deal residual activities by **age based withdrawal**

Withdraw all content after a preset time T (e.g. 24 hours)

Snapchat, Cyber dust

Both original post as well as residual activities are withdrawn

Dealing with residual activities is difficult

Straw man:

Withdraw all the residual activities with original tweet/account

Problem

Residual activities are **not "owned"** by the original poster

Emerging OSMs deal residual activities by **age based withdrawal**

Withdraw all content after a preset time T (e.g. 24 hours)

Snapchat, Cyber dust

Both original post as well as residual activities are withdrawn

Problem with age based withdrawal

1. Do **not facilitate interaction** with content
2. **No archive** of past activities, **no long term memory**

Dealing with residual activities is difficult

Straw man:

Withdraw all the residual activities with original tweet/account

Problem

Residual activities are **not "owned"** by the original poster

Emerging OSMs deal residual activities by **age based withdrawal**

Withdraw all content after a preset time T (e.g. 24 hours)

Snapchat, Cyber dust

Both original post as well as residual activities are withdrawn

Problem with age based withdrawal

1. Do **not facilitate interaction** with content ☑
2. **No archive** of past activities, **no long term memory**

One idea: Inactivity based withdrawal

Automatically withdraw content only when it is inactive

Inactive content: **no interaction** (e.g., retweets) for **time T**

In other words

No tweet receiving **recent interaction** will be withdrawn

One idea: Inactivity based withdrawal

Automatically withdraw content only when it is inactive

Inactive content: **no interaction** (e.g., retweets) for **time T**

In other words

No tweet receiving **recent interaction** will be withdrawn

Through simulation we discovered

Inactivity based withdrawal **allows more interactions**

Especially for popular tweets -- Details in the paper

One idea: Inactivity based withdrawal

Automatically withdraw content only when it is inactive

Inactive content: **no interaction** (e.g., retweets) for **time T**

In other words

No tweet receiving **recent interaction** will be withdrawn

Through simulation we discovered

Inactivity based withdrawal **allows more interactions**

Especially for popular tweets -- Details in the paper

Even for this idea **no archive** of past activities, **no long term memory**

Summary

Analyzed longitudinal exposure control from recent to very far past

- Users control exposure by withdrawing **surprisingly large amount of old data**

First study to analyze information leakage via **residual activities**

- They **leak significant information** about **withdrawn content**

Inactivity based withdrawal is an approach **to stop information leakage** from residual activities and **facilitate interaction**

THANKS!

Check out our Twitter web app to see your information leakage via residual activities: http://twitter-app.mpi-sws.org/footprint/