Privacy, intellectual property and censorship Issues in OSNs

Panagiotis Ilia

<pilia@ics.forth.gr>

Foundation for Research and Technology – Hellas (FORTH)

Outline

Part 1: User Privacy – Content Ownership

- Introduction
- Proposed Mechanism
- Evaluation / Limitations
- Intellectual Property

Part 2: Censorship in OSNs

- Motivation
- Objectives
- Analysis

Part 1:

User Privacy
&
Content Ownership

Introduction

- Online Social Networks -

Popularity of OSNs

- **Facebook**: > **1.5b** monthly active users
- **Twitter**: > **320m** monthly active users

Uploaded content

- **Facebook**: > **350m** photos uploaded daily
- **Twitter**: > **500m** tweets sent daily

Introduction
- User Privacy -

Photos uploaded online

- Contain users' personally identifiable information (PII)
- Reveal private information (e.g., relationships, locations)

Users disclose sensitive information

- No concern about privacy
- Unaware of implications / consequences
- Unaware of true visibility of shared content

Motivation

- Group Photos -

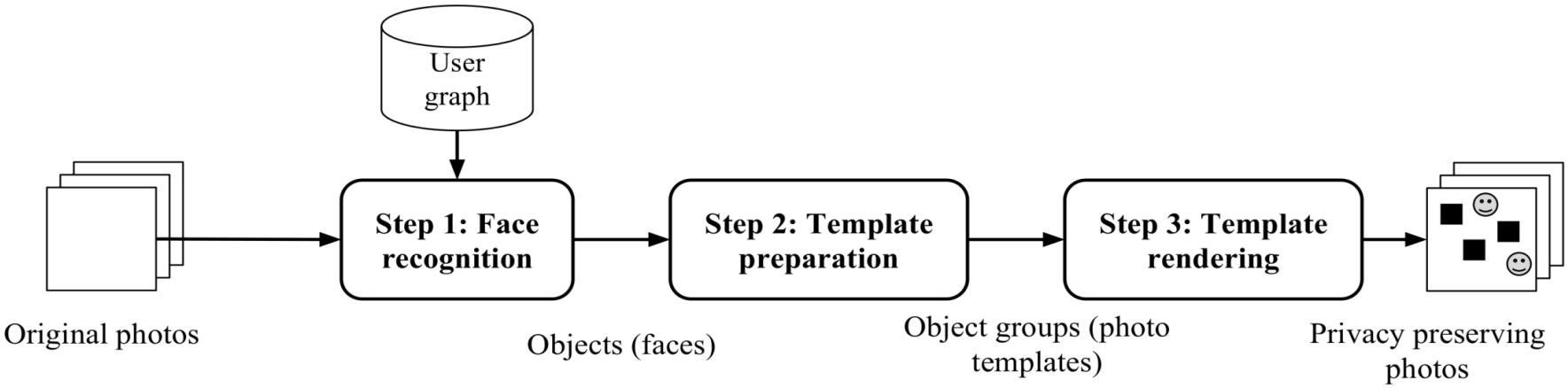
Usually photos depict multiple individuals

- Users cannot **control** data published by others
- The **uploader** is considered **owner** of the photo.

  → Granted full rights on the photo.
- Depicted/tagged people are **NOT** considered co-owners.

  → **Cannot** restrict access or remove it

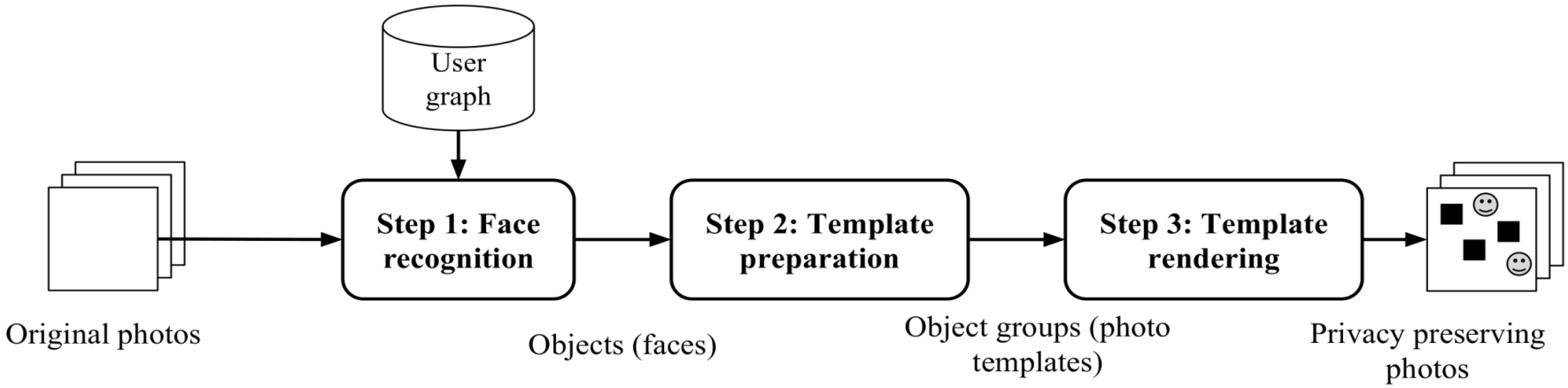
System Design



Allow each user in photo to control disclosure of PII

- Changes **granularity** of AC from **photo** to users' **faces**.

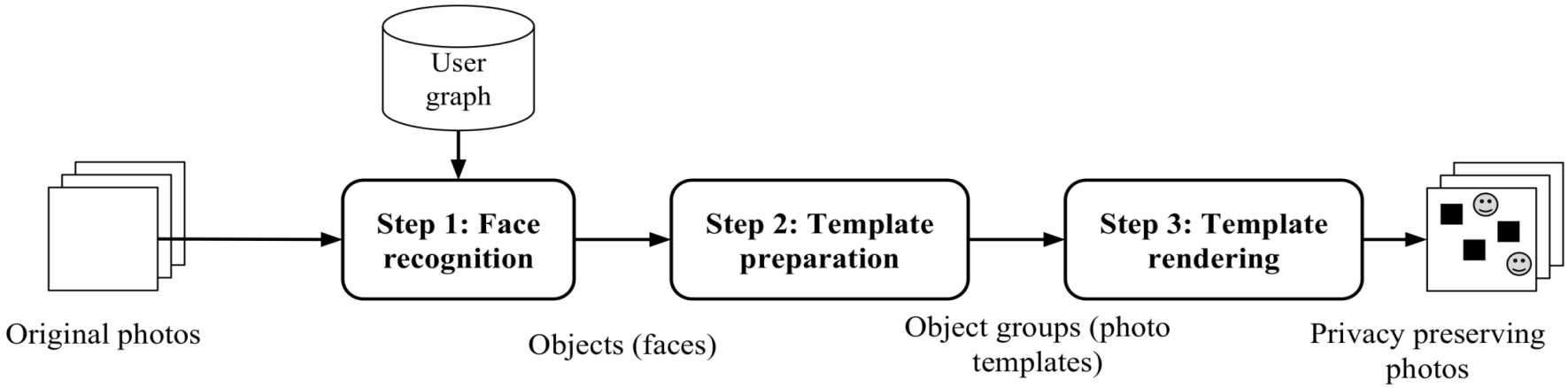
System Design



Step 1: Face Recognition

- Exploit social relationships
- Uploader's friends, friends of identified users etc.

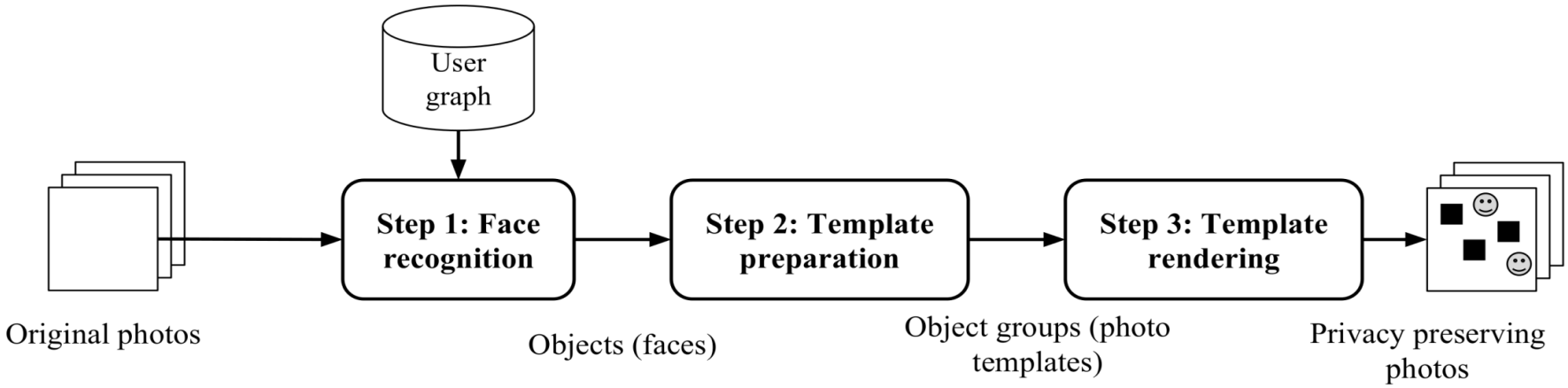
System Design



Step 2: Template Preparation

- N transparent layers, each contains a single blurred face
- Each identified user sets its own permissions

System Design



Step 3: Template Rendering

- Determines which faces the accessing user has permissions to view
- Generates a "**processed**" photo "**on the fly**"

Evaluation

- Privacy / Effectiveness -

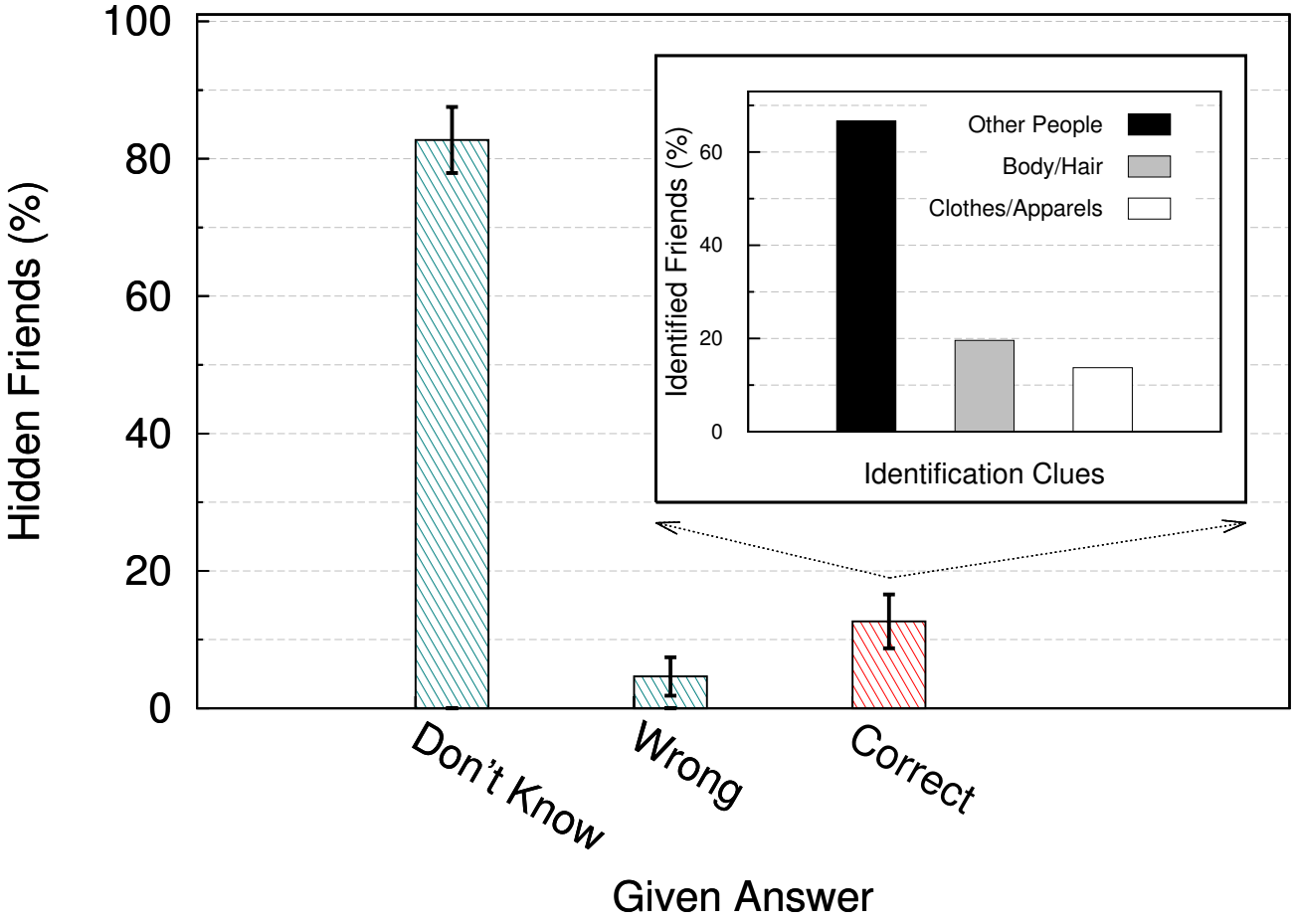
User Study

- 34 participants
- 14 challenges per participant

- One friend “hidden” in each challenge
- Requested to identify the “hidden” friend

Evaluation

- Privacy / Effectiveness -



Why we propose this mechanism

- It can significantly enhance users' privacy on shared content
- It can be easily deployed by current OSNs, as an additional module
- Small overhead / Scalable

Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi and Sotiris Ioannidis. **Face/Off: Preventing Privacy Leakage From Photos in Social Networks.** In Proceedings of the 22nd ACM Conference on Computer and Communications Security (**CCS '15**).

But

"Malicious" users befriend victim and its friends

- Collect information/photos
  - Impersonation attack
  - Bypassing social authentication mechanism [1]
- Can download and re-upload photos of others

The proposed approach relies on **face recognition**

- Difficult to identify strangers (multi-hop friends)
- Photos may not depict any face

[1] Iasonas Polakis, Panagiotis Ilia, Federico Maggi, Marco Lancini, Georgios Kontaxis, Stefano Zanero, Sotiris Ioannidis, Angelos D. Keromytis. **Faces in the Distorting Mirror: Revisiting Photo-based Social Authentication** In Proceedings of the 2014 ACM Conference on Computer and Communications Security **(CCS '14)**

Ownership - Intellectual Property
- Proposed Approach -

The OSN implements a credit system

Each user sets its rules and required amount of credits.

A user can access the photo after transferring the credits

The OSN identify the uniqueness of each uploaded photo

- Watermarking
- Fingerprinting

When a photo is **re-uploaded**, the correct rules and credits are applied

Intellectual Property
- Challenges -

The Watermarking and Fingerprinting algorithms should be:

- Resistant to OSN transformations (resizing, cropping, compression)
- Non detectable by users
- Tamperproof / non-reversible

The system should be:

- Efficient (low overhead)
- Scalable

Part 2:

Censorship in OSNs

Censorship in Twitter

- "Country Withheld Content" policy -

From January 2012

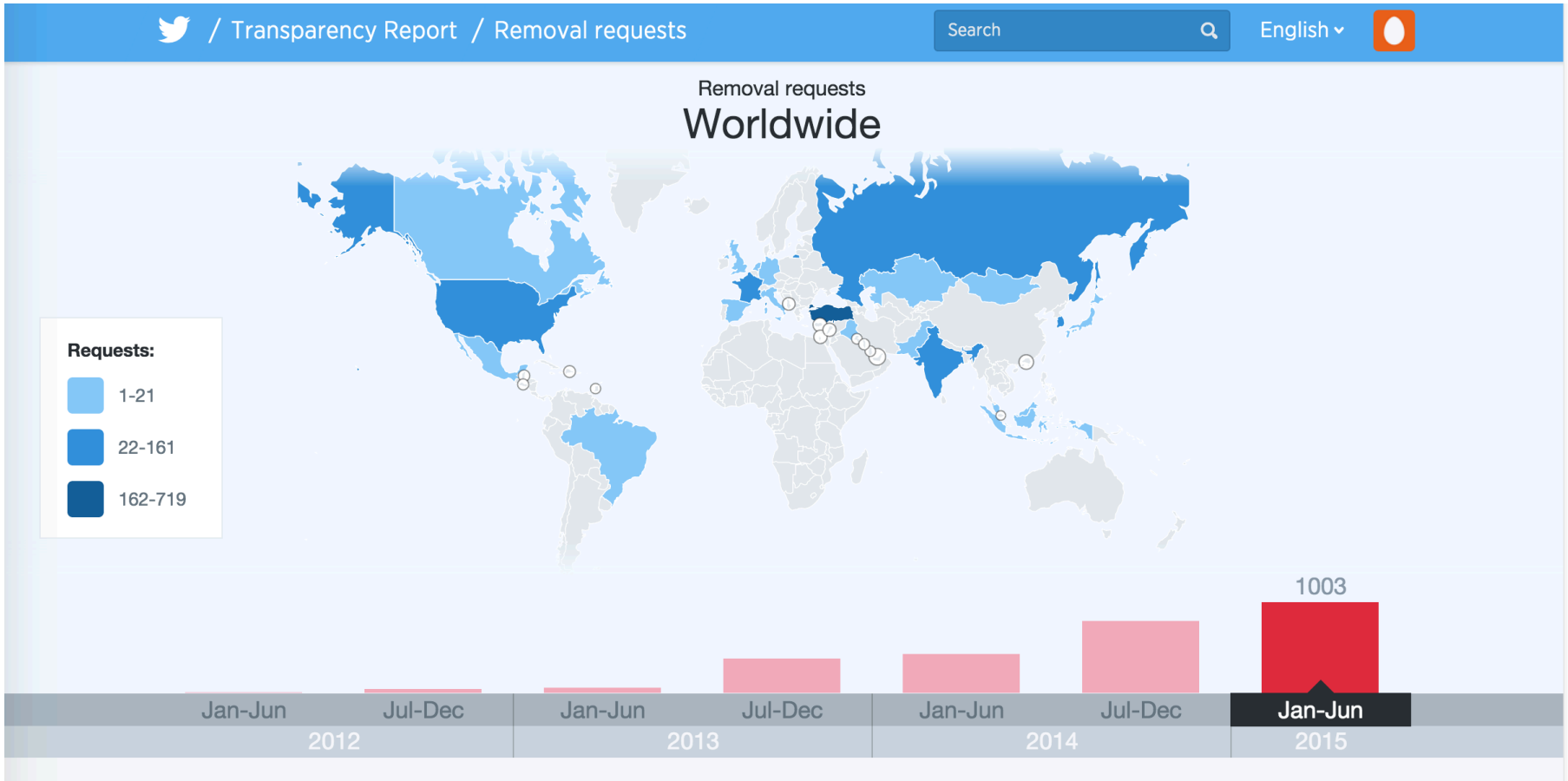
- Governments / law enforcement agencies can request Twitter to withhold content / accounts.
- Twitter checks if this content violates its "Terms of Use"
- Those requests are published on lumendatabase.org
- Twitter publishes transparency reports

Censorship in Twitter

- "Country Withheld Content" -

Censorship in Twitter

- Transparency Reports -



Censorship in Twitter
- Objectives -

- Comparison of withheld and non-withheld tweets
- Investigate how tweets are being chosen for being withheld
- Graph properties, similarities and differences, clusters
- Investigate if there are patterns in user behaviour
- Influence of these users and propagation of tweets

Analysis

- Tweets in Dataset -

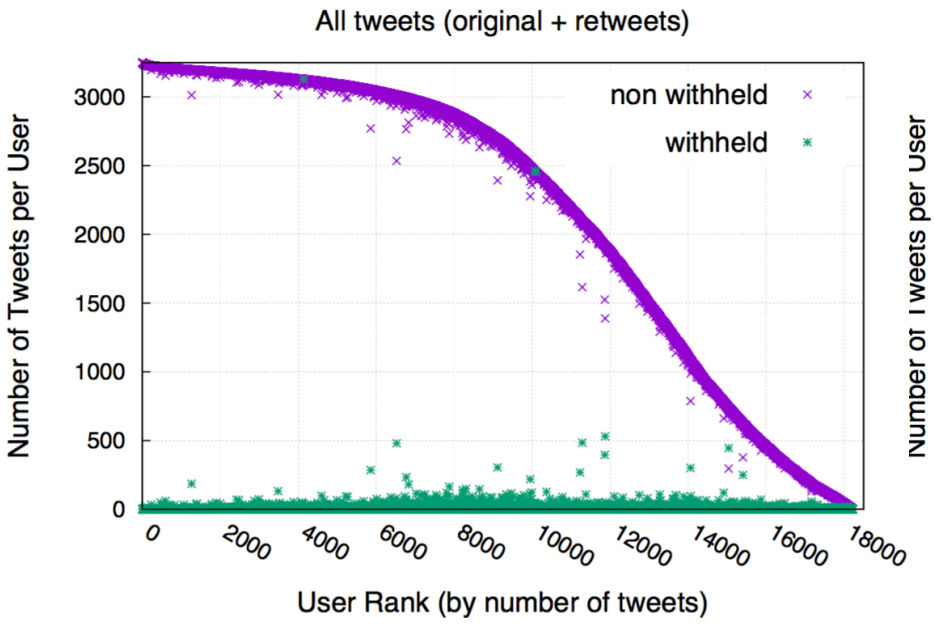
18225 users

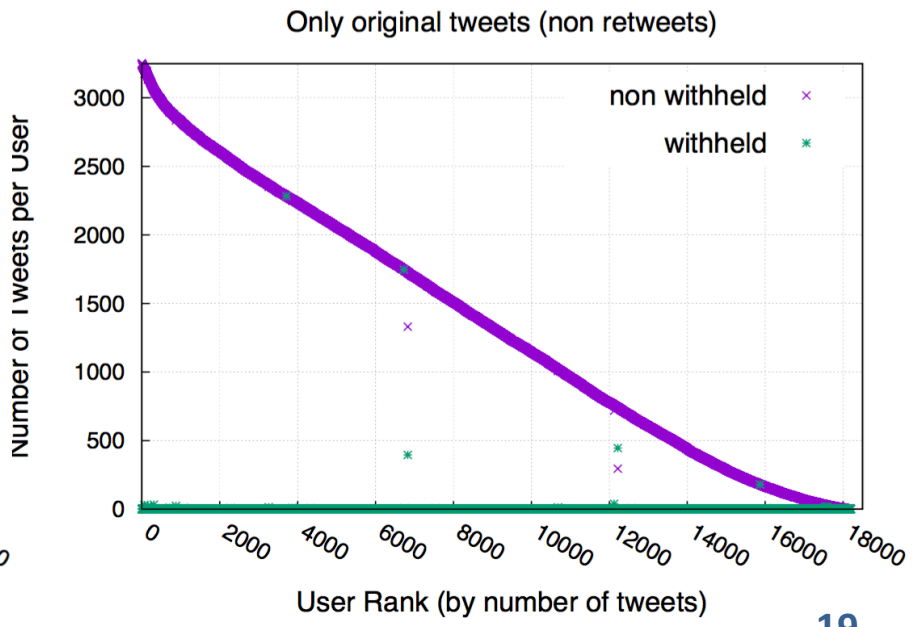
~ 39m tweets

~ 77k withheld tweets

~ 24.5m tweets

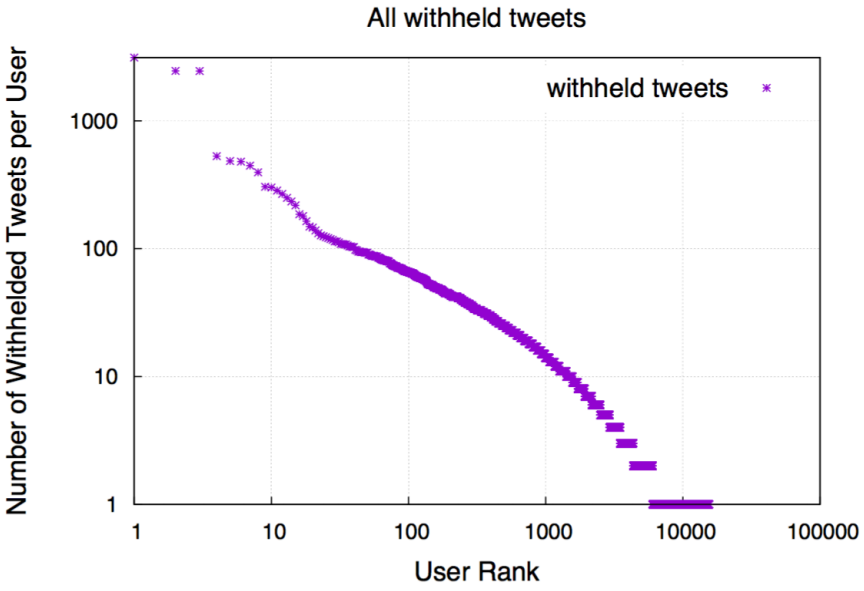
~ 7186 withheld tweets

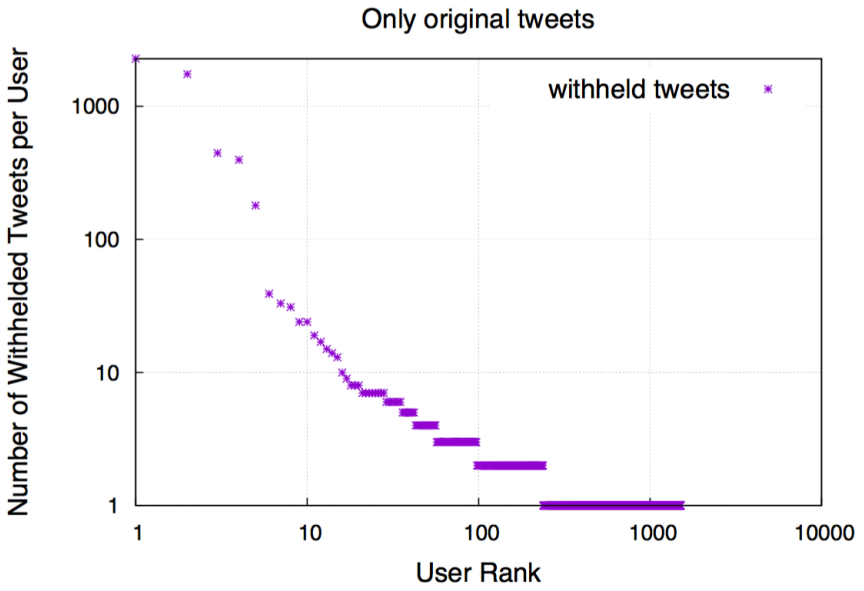




Analysis

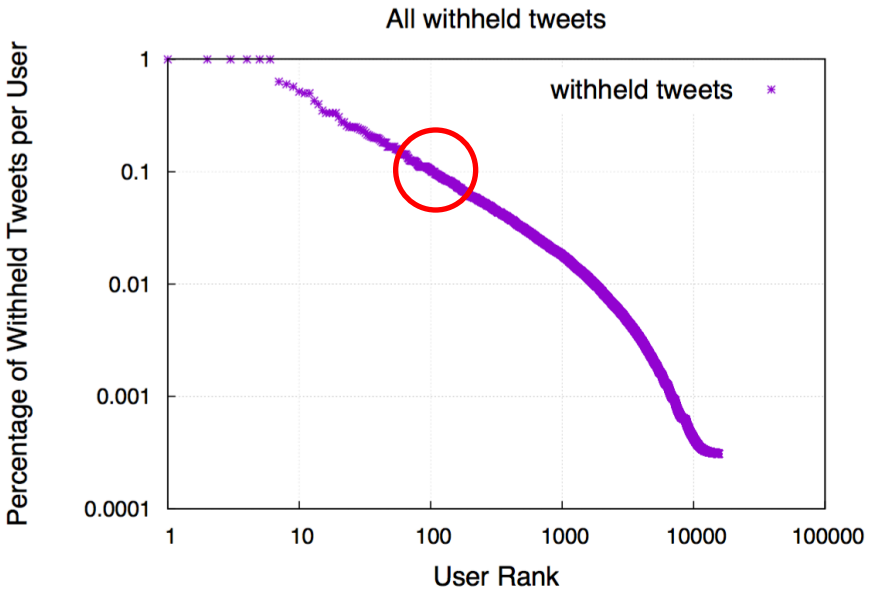
- Withheld tweets per user -

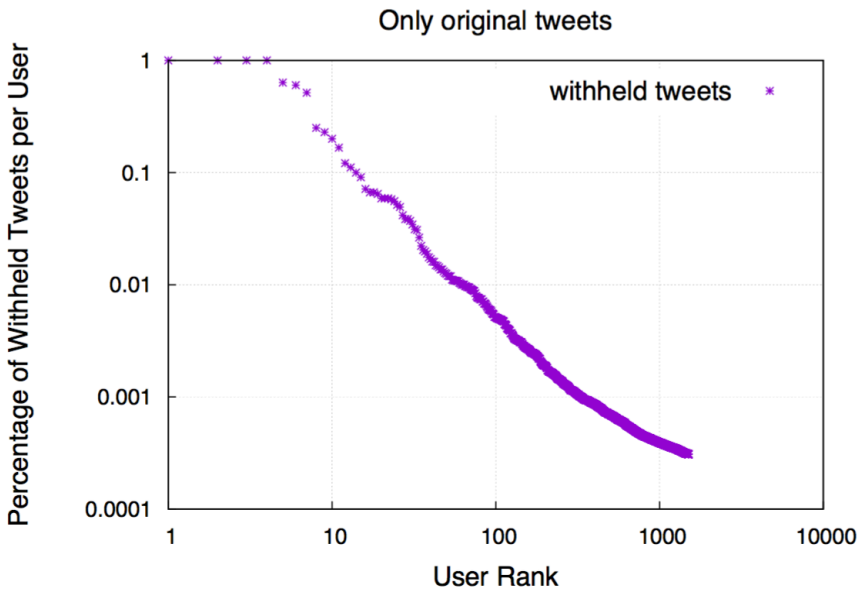




Analysis

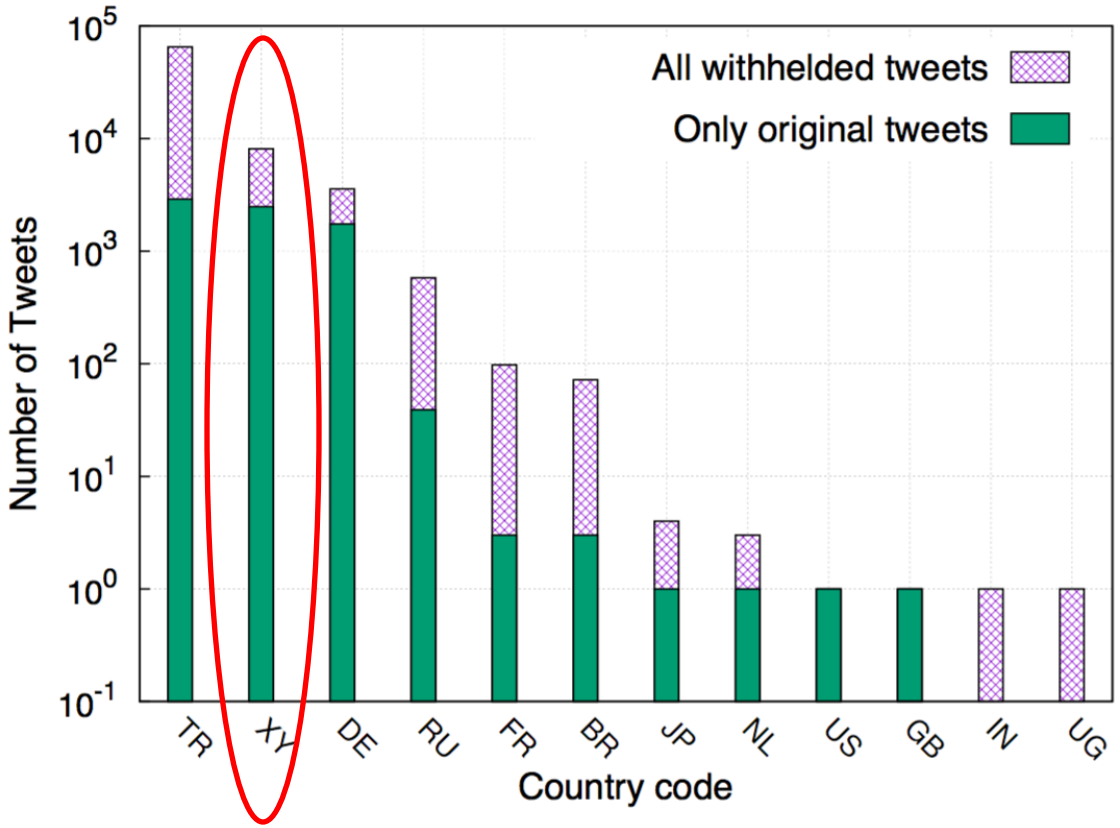
- Percentage of Withheld tweets per user -





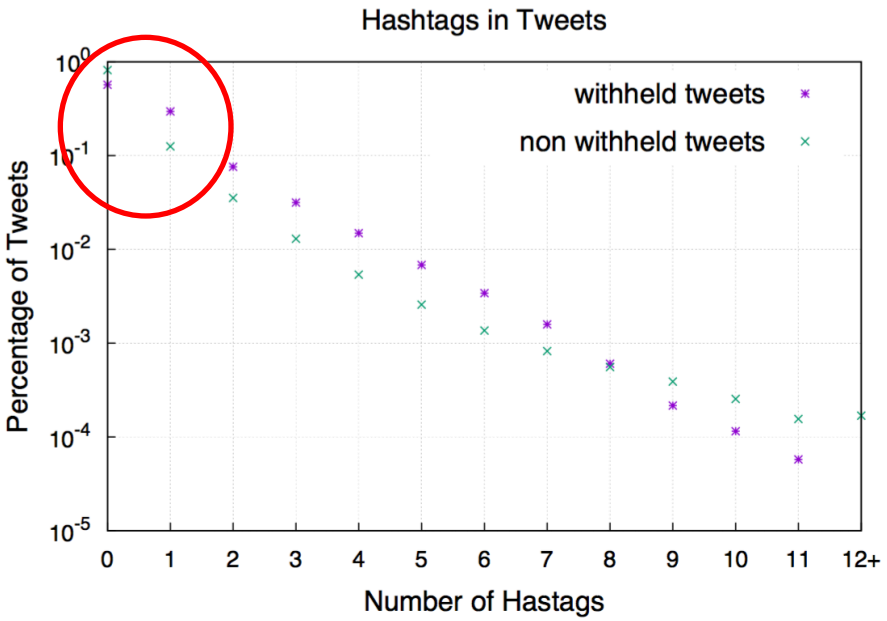
Analysis

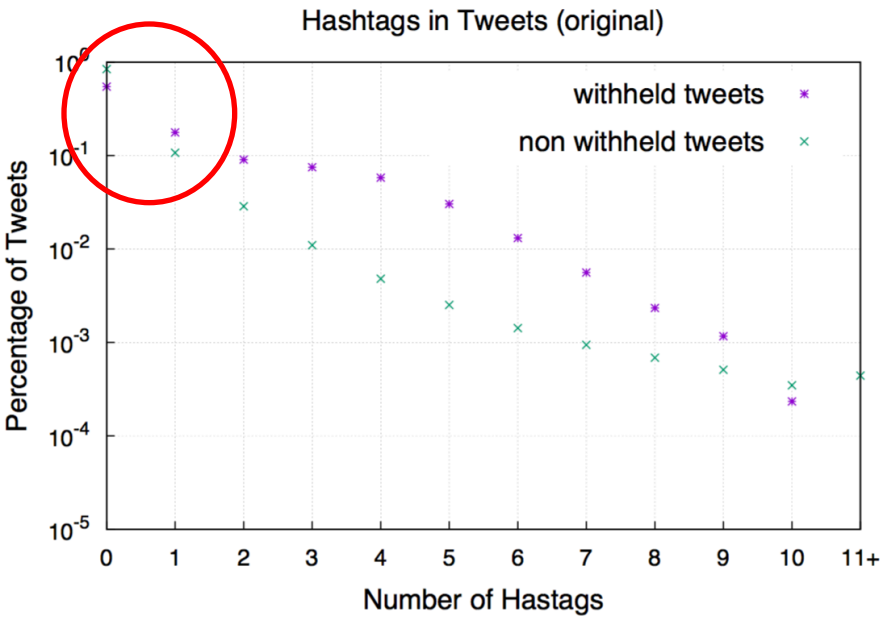
- Withheld tweets per Country -



Analysis

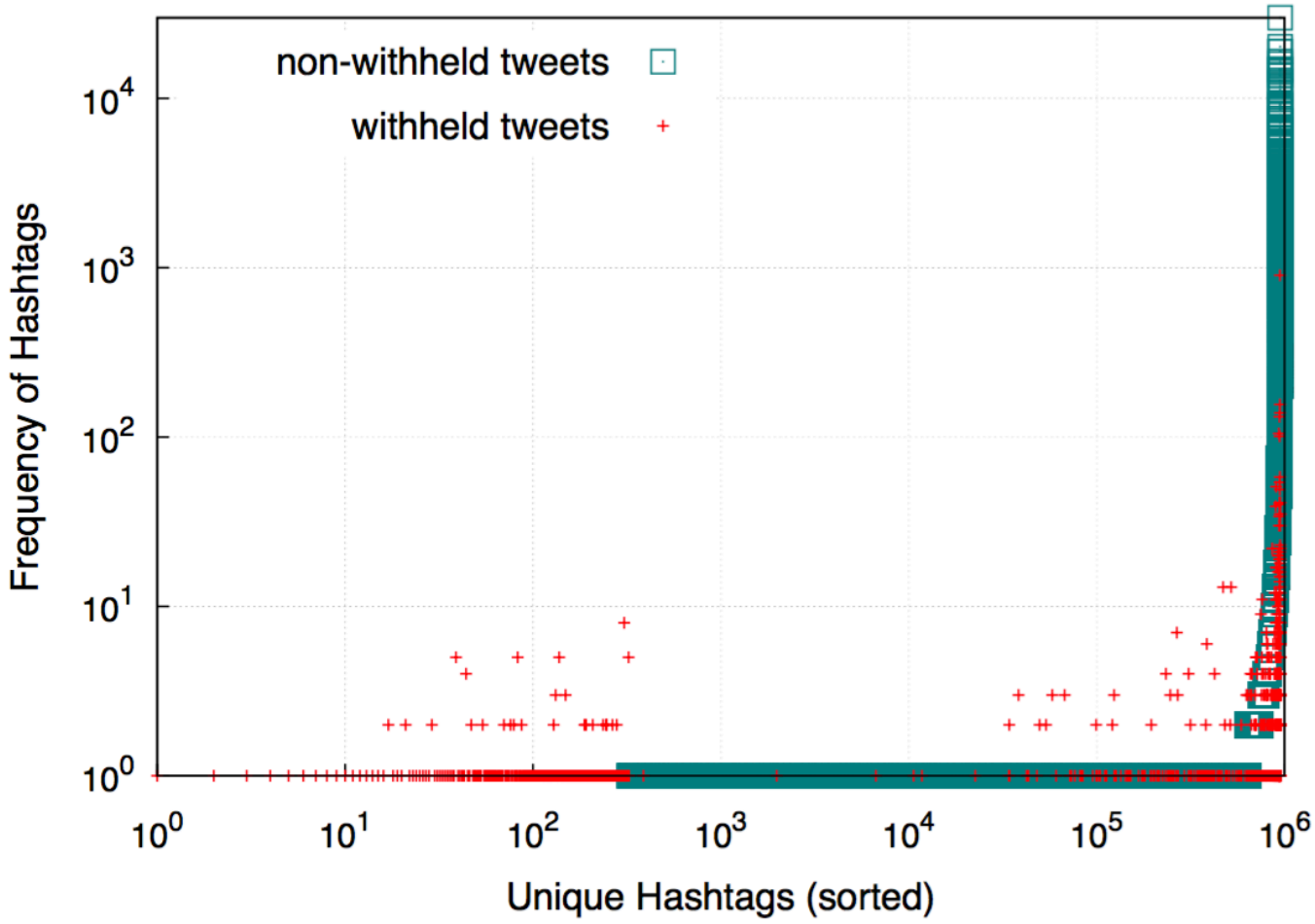
- Number of Hashtags in Tweets -





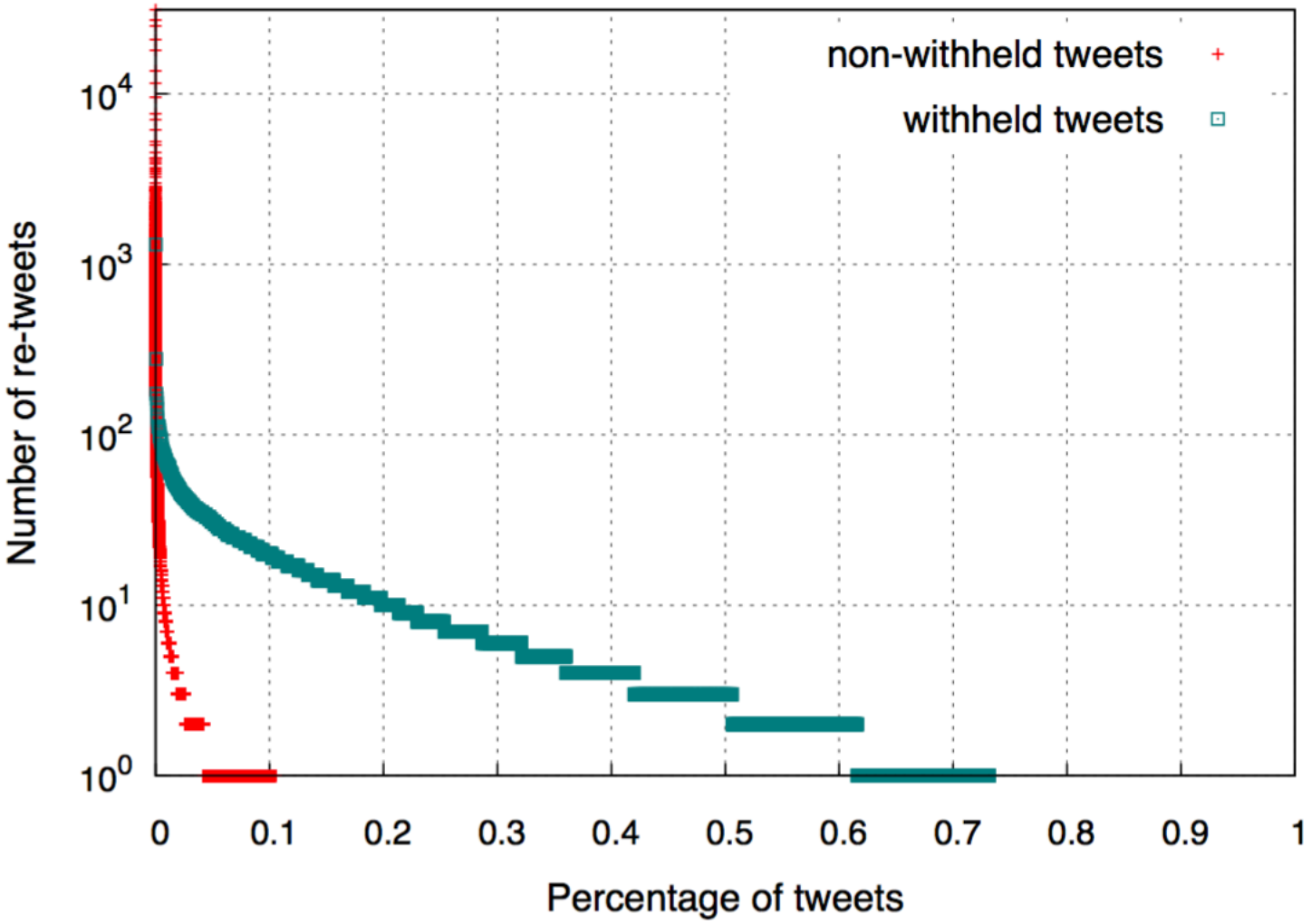
Analysis

- Keywords used as Hashtags -



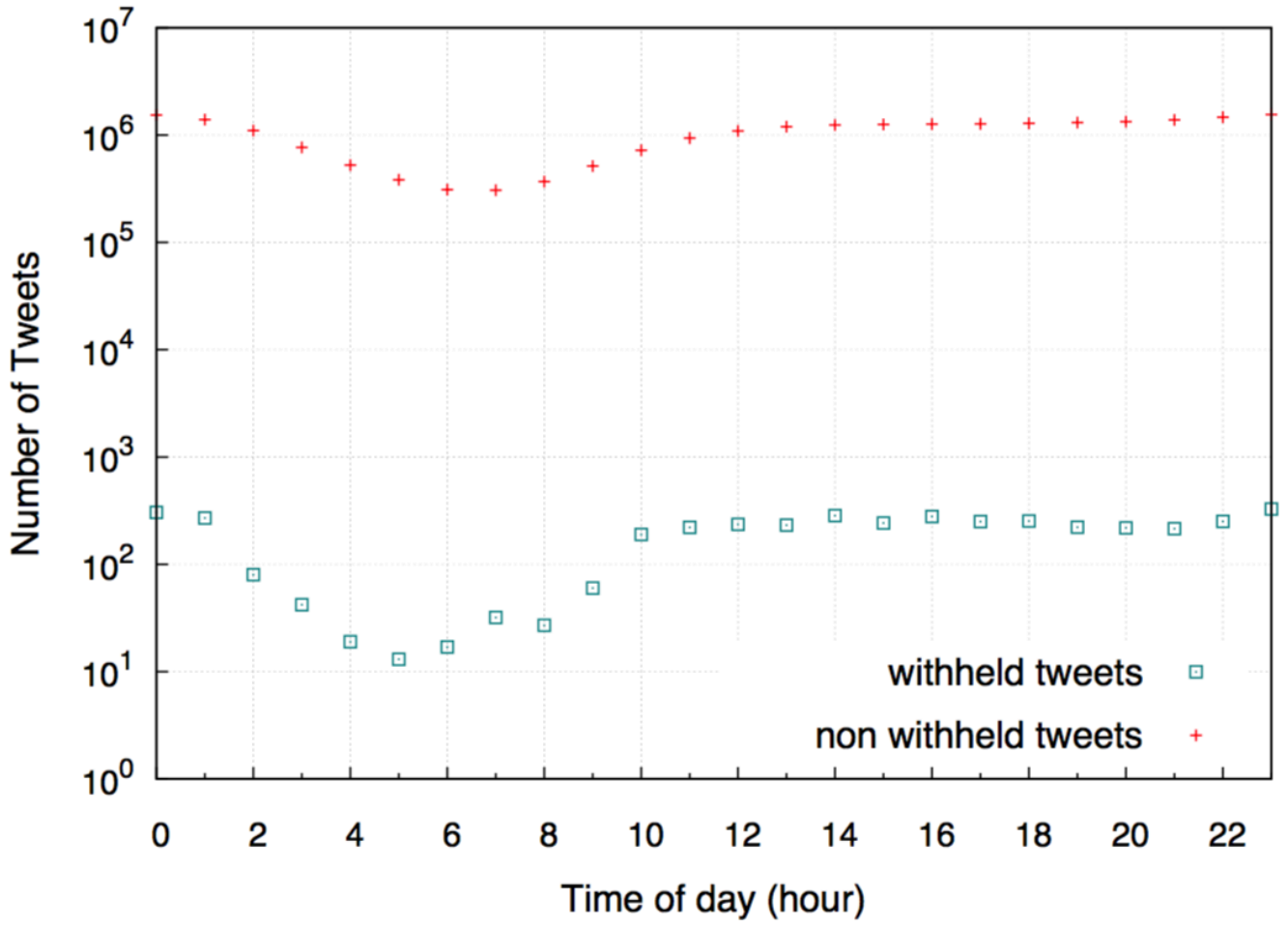
Analysis

- Retweets per Original Tweet -



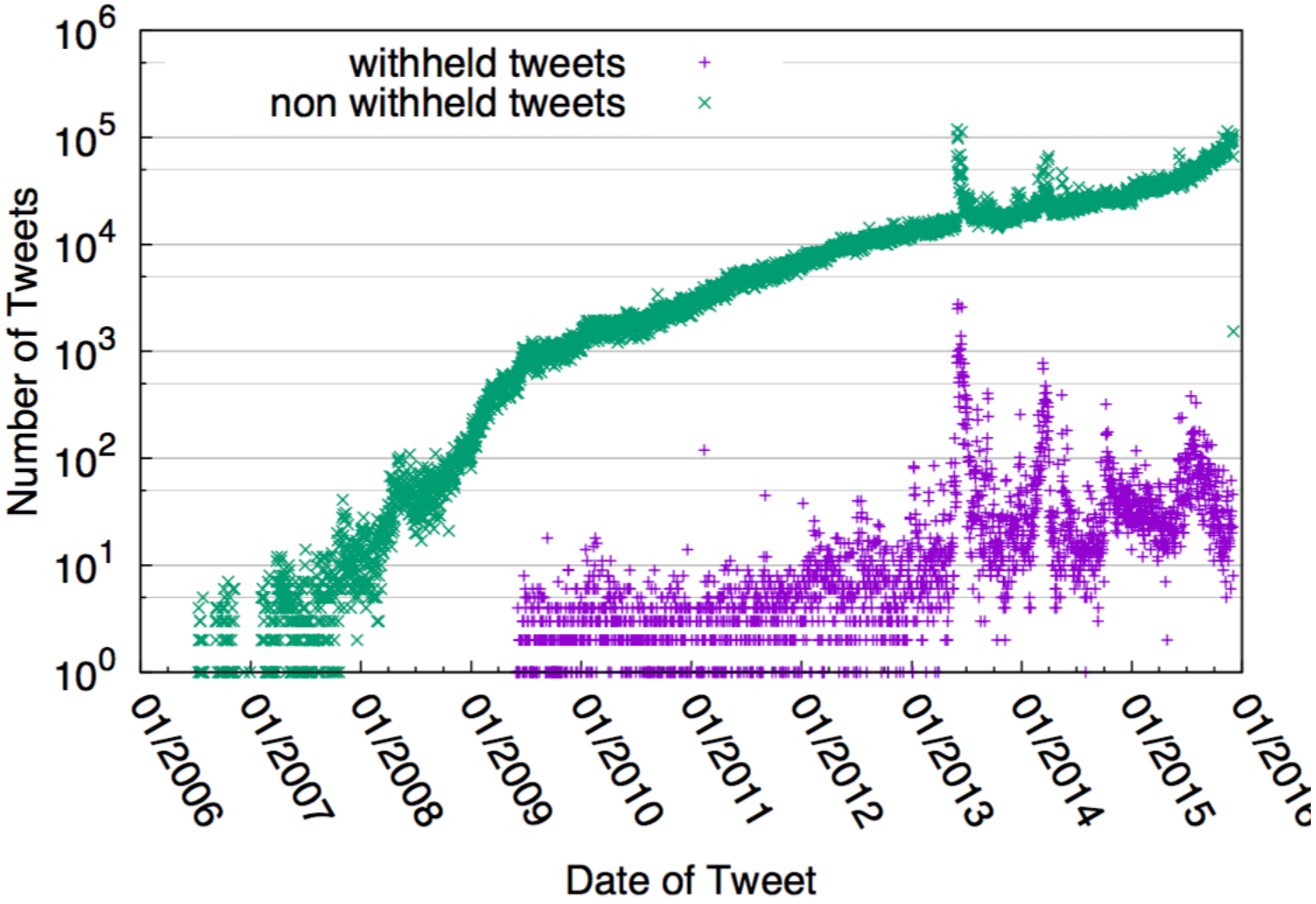
Analysis

- Tweets per Time of Day -



Analysis

- Tweet date (creation) -



Future Work

- We collect a new dataset (original tweets of re-tweets)
- Estimate how long tweets live before being withheld
- Investigate connections between users that have withheld tweets
- Identify Influence of these users and propagation of tweets

Summary

- Design a fine-grained access control mechanism
  - Small overhead, scalable, effective

- Design a new model for solving intellectual property issues
  - Resistant fingerprinting and watermarking algorithms.

- Study cases of censorship in Tweeter
  - Compare withheld to non-withheld tweets
  - Graph clustering for users with withheld tweets
  - Propagation of withheld tweets, despite withholding efforts

Thank You