

# Fifteen Minutes of Unwanted Fame: Detecting and Characterizing Doxing

Peter Snyder\* – Periwinkle Doerfler+ – Chris Kanich\* – Damon McCoy+



# Overview

- Doxing is a targeted form of online abuse
- Prior work is qualitative or on defensive techniques
- We don't understand the scale or targets of problem
- This work is the first quantitative, large scale measurement of doxing

# Outline

- Problem area
- Measurement methodology
- Results and findings
- Discussion and conclusions

# Outline

- Problem area
- Measurement methodology
- Results and findings
- Discussion and conclusions

# What is Doxing? (1/2)

- Method of targeted online abuse
- Attackers compile sensitive information about the target
  - **Personal:** Name, addresses, age, photographs, SSN
  - **Relationships:** Family members, partners, friends
  - **Financial:** Work history, investments, CCN
  - **Online:** Email, social network accounts, passwords, IPs

# What is Doxing? (2/2)

- Information is compiled into plain text files
- Released "anonymously"
  - Text sharing sites (e.x. [pastebin.com](https://pastebin.com), [skidpaste.com](https://skidpaste.com))
  - Online forums (e.x. 4chan, 8chan)
  - Torrents
  - IRC, Twitch, social networks, etc.

=====  
Full Name: [REDACTED] [REDACTED]

Aliases> [REDACTED]

Age: [REDACTED]

DOB: [REDACTED]/[REDACTED]/[REDACTED]

Address: [REDACTED] [REDACTED] [REDACTED] [REDACTED], [REDACTED] [REDACTED] // Confirmed

Mobile Number: +[REDACTED] ([REDACTED]) [REDACTED]-[REDACTED] // Confirmed

Email: [REDACTED]@[REDACTED].[REDACTED] // Confirmed

Illness: Asthma

=====  
ISP Records>

ISP: Rogers Cable // Previous

IP Address: [REDACTED].[REDACTED].[REDACTED].[REDACTED] // Previous  
=====

Parental Information>

Father: [REDACTED] [REDACTED] [REDACTED]

Age: [REDACTED]

Aliases) [REDACTED], [REDACTED], [REDACTED]

Name) [REDACTED] [REDACTED]

DOB [REDACTED]/[REDACTED]/[REDACTED]

Address) [REDACTED] [REDACTED] [REDACTED], [REDACTED], [REDACTED]

Cell Phone) [REDACTED]-[REDACTED]-[REDACTED] - Sprint, Mobile

Caller ID) [REDACTED] [REDACTED]

Old Home Phone) [REDACTED]-[REDACTED]-[REDACTED] - CenturyLink, Landline

Last 4 of Mastercard) [REDACTED]

Emails) [REDACTED]@[REDACTED].[REDACTED], [REDACTED]@[REDACTED].[REDACTED]

Snapchat) [REDACTED]

Twitter) @[REDACTED]

Facebook) [https://facebook.com/\[REDACTED\]](https://facebook.com/[REDACTED]), [REDACTED]

Skype) [REDACTED], [REDACTED]



# Doxing Harms



# Frequency, Targets and Effects

- Prior work is based in qualitative or preventative / risk management approaches
- Research Questions:
  1. How frequently does doxing happen?
  2. What information is shared in doxes? Who is targeted?
  3. What is knowable about the large scale effects and harms?
  4. Are anti-abuse tools effective?

# Outline

- Problem area
- Measurement methodology
- Results and findings
- Discussion and conclusions

# Steps to Protect Victims

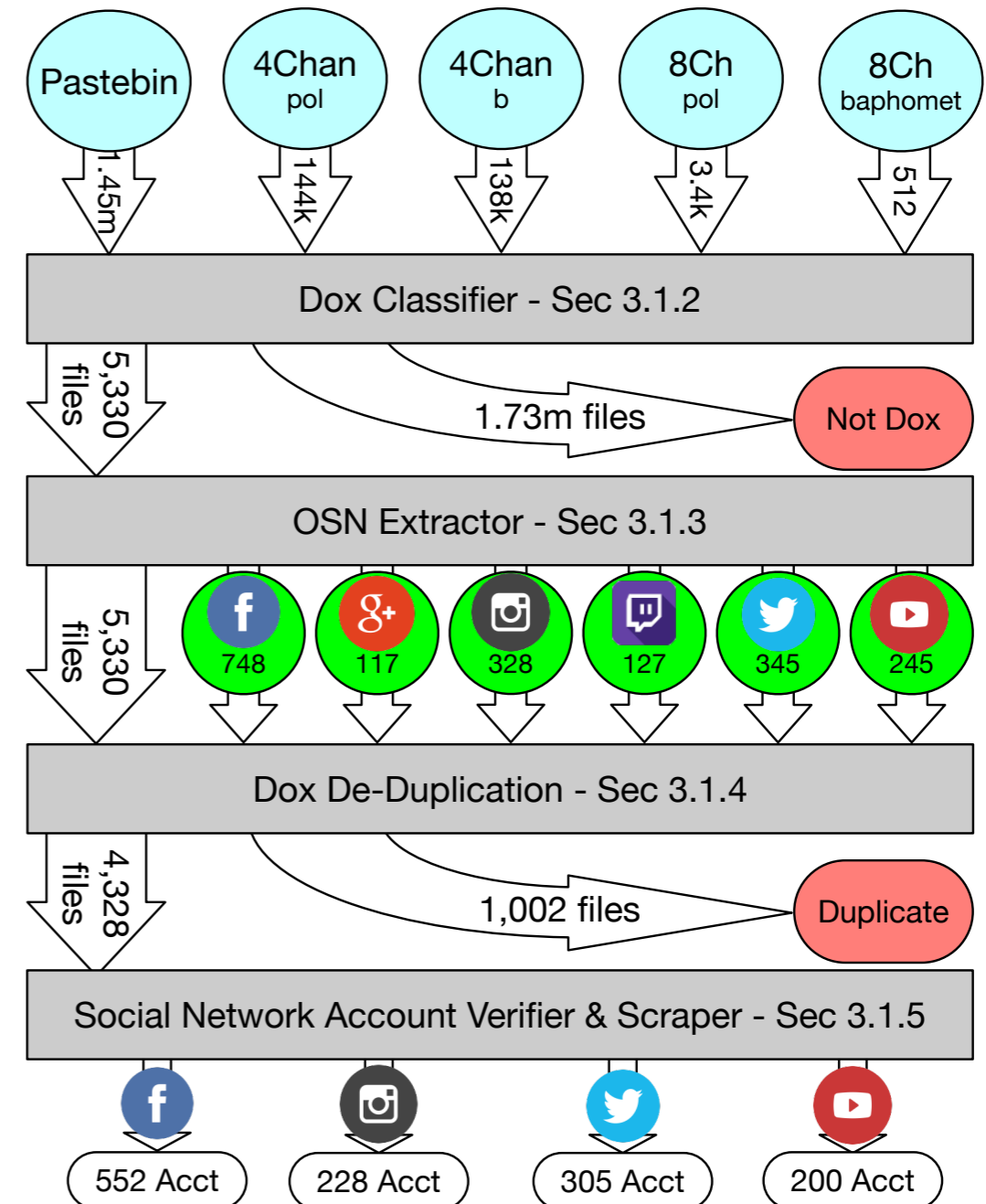
- Worked closely with IRBs; multiple rounds of study design
- Only recorded publicly available data, careful to not use it to record data
- Careful data storage / analysis methods: only recorded high level summary data
- Data protection best practices (key based encryption, single data store, strict access controls)

# General Measurement Strategy

- Find places online where doxes are frequently shared
- Train a classifier to determine how much activity is doxing
- Measure extracted doxes to determine contained information
- Watch the OSN accounts of doxing victims for abuse

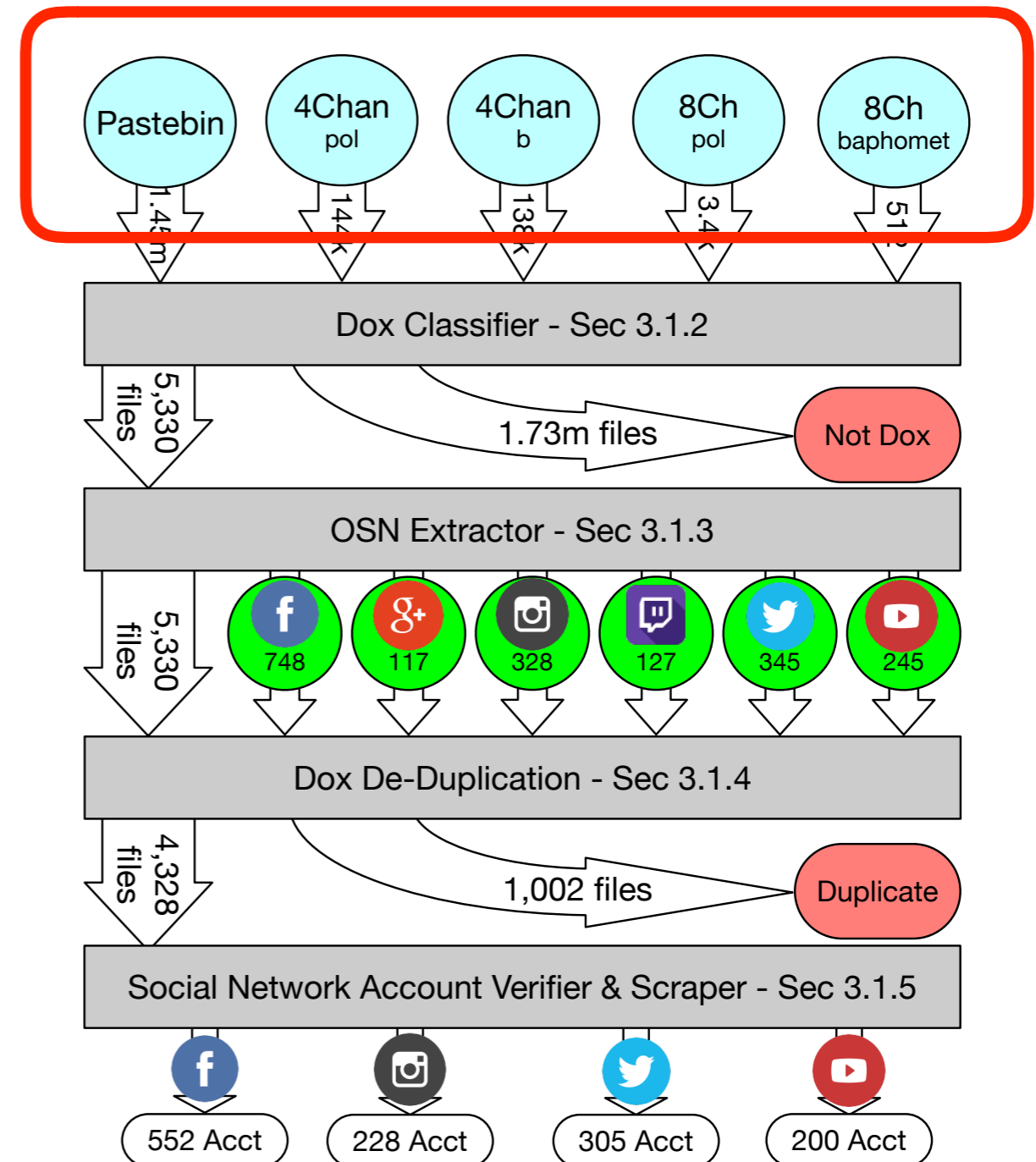
# Dox Collection Pipeline

- Fully automated
- Single IP at the University of Illinois at Chicago
- Two recording periods:
  - Summer of 2016
  - Winter of 2016



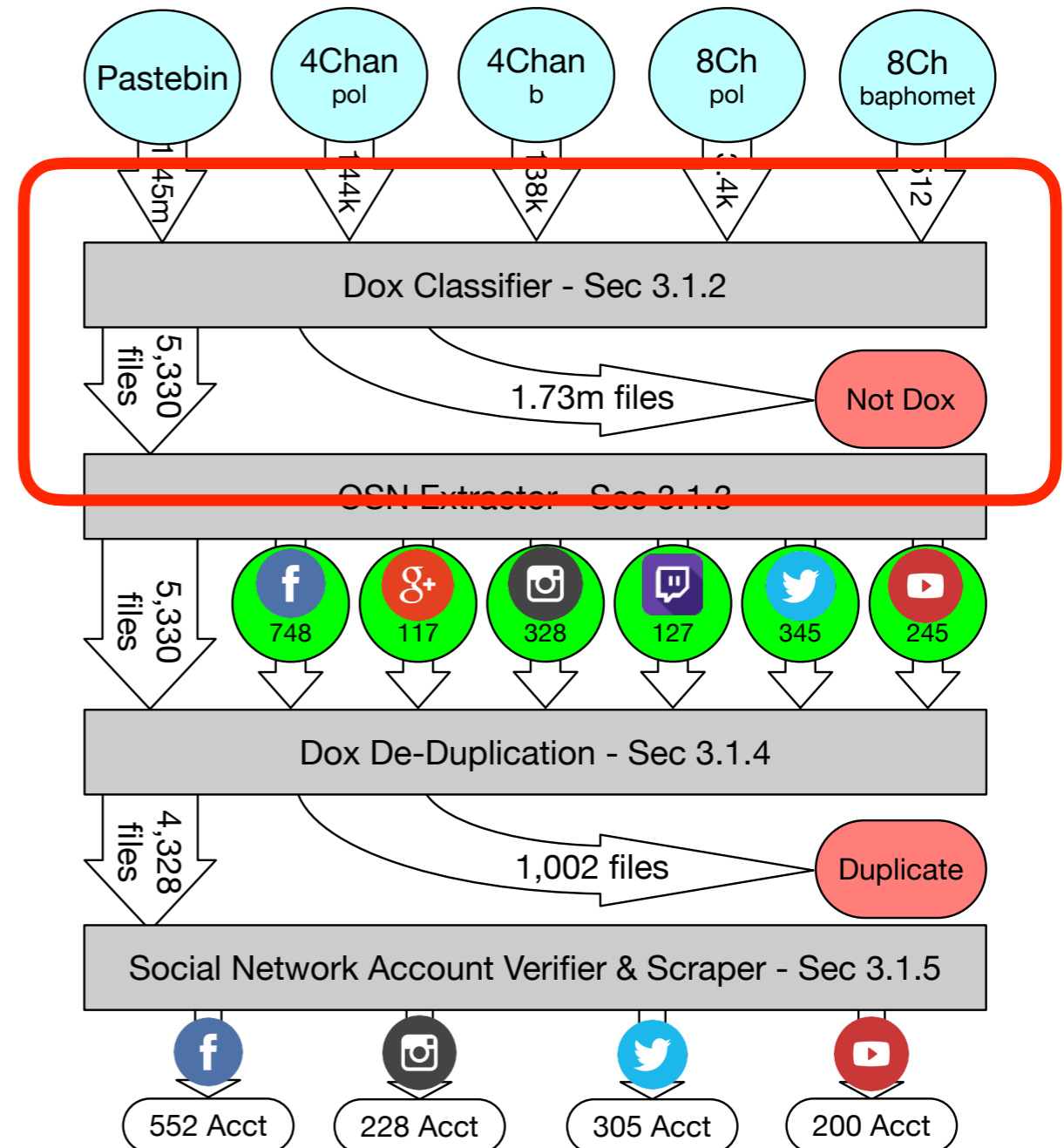
# Text File Collection

- Data recorded from
  - [pastebin.com](https://pastebin.com)
  - [4chan.org](https://4chan.org) (pol, b)
  - [8ch.net](https://8ch.net) (pol, baphomet)
- Selected because:
  - "Original" sources of doxes
  - Anecdotal reputation for doxing



# Text File Classification

- Scikit-learn, TfidfVectorizer, SGDClassifier
- Training Data:
  - Manual labeling of Pastebin crawl
  - "proof-of-work" sets



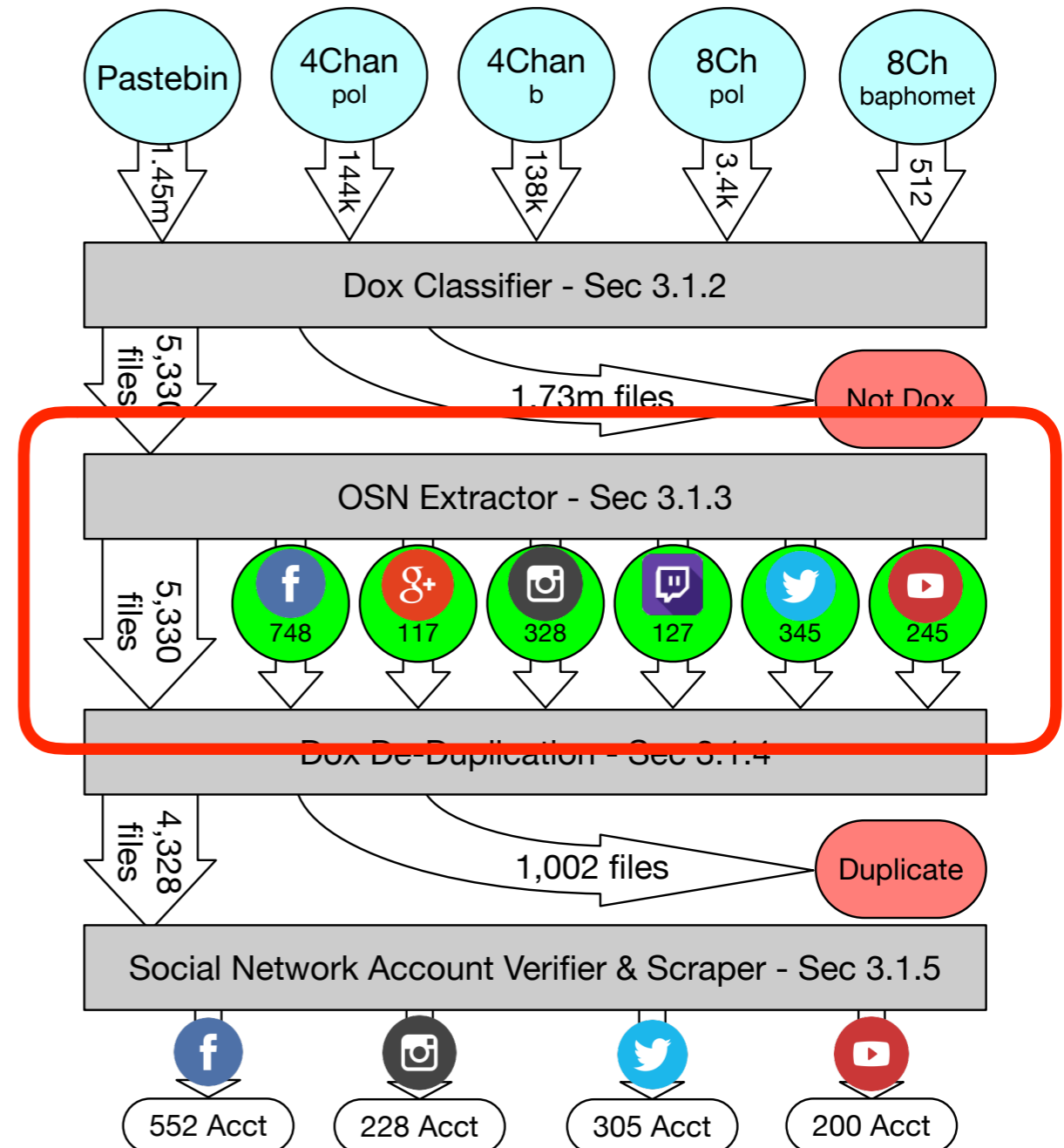


# Text File Classification

Label	Precision	Recall	# Samples
<b>Dox</b>	0.81	0.89	258
<b>Not</b>	0.99	0.98	3,546
Avg / Total	0.98	0.98	3,804

# Social Networking Account Extractor

- Extract social networking accounts
- Custom, heuristic-based identifier
- Evaluated on 125 labeled doxes

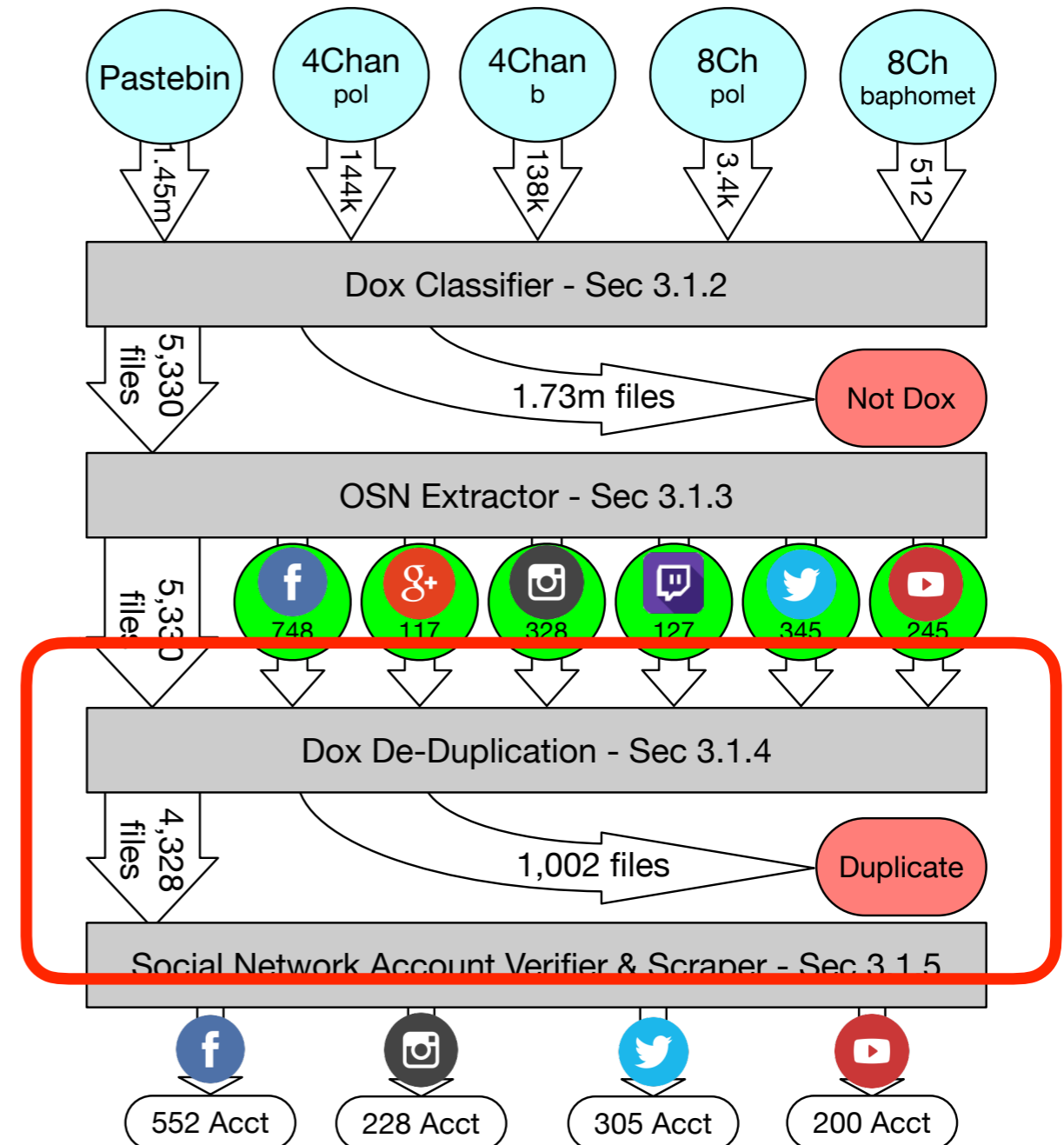


# Social Networking Account Extractor

	% Doxes Including	Extractor Accuracy
Instagram	11.2	95.2
Twitch	9.7	95.2
Google+	18.4	90.4
Twitter	34.4	86.4
Facebook	48.0	84.8
YouTube	40.0	80.0

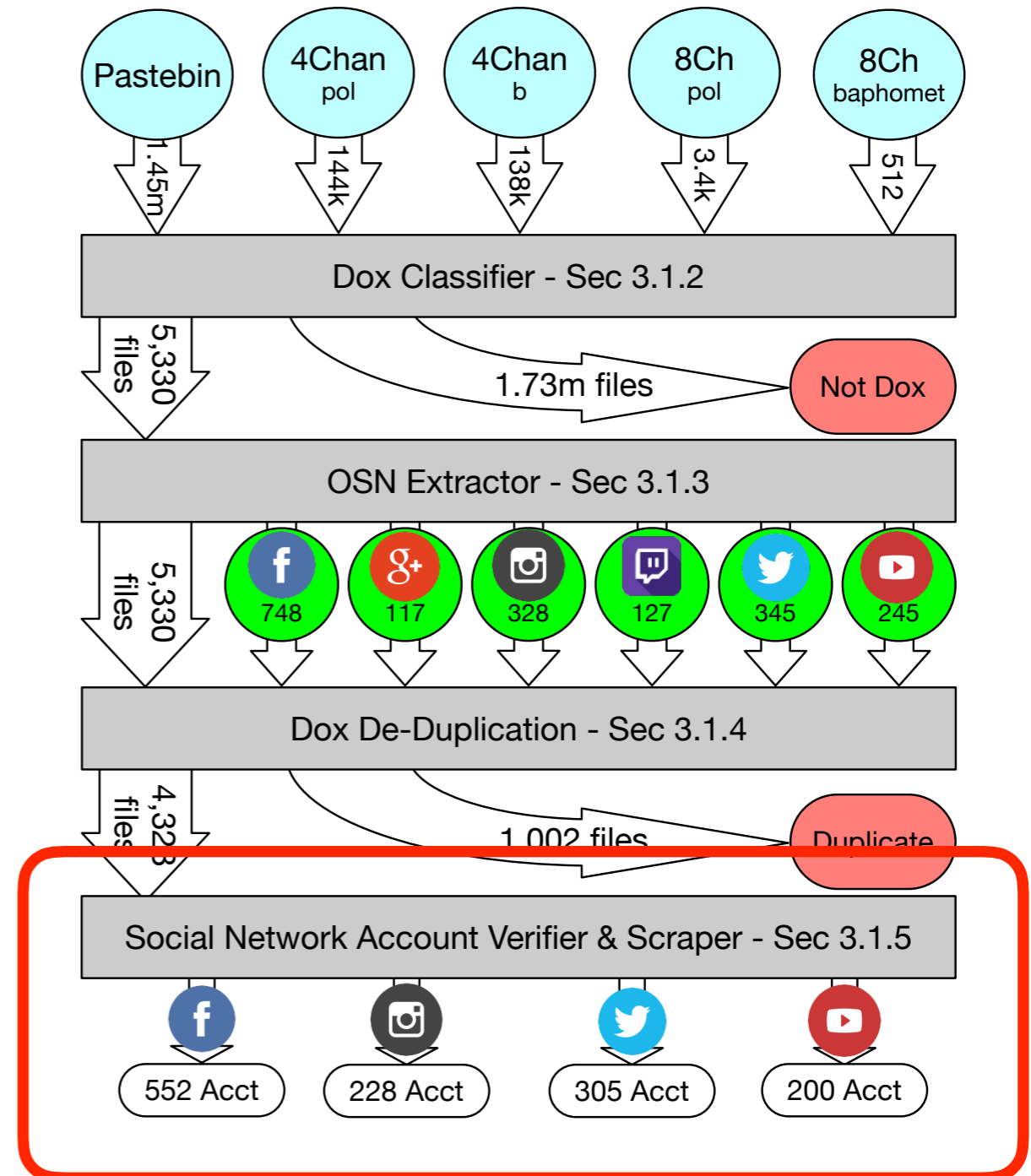
# Dox De-duplication

- Similar doxes, identical target
- Hash based comparison fragile to marginal updates
- Compare referenced OSN accounts
- ~14.2% of doxes were duplicates



# Social Network Status Watcher

- Repeatedly visit referenced OSN accounts
- After 1, 2, 3, 7, 14... days
- Only record the status of the account:
  - public, private, inactive
- Single IP @ UIC



# Manual Dox Labeling

- Randomly selected 464 doxes
- Manually label each dox to understand the contents.
  - Did it include name, address, phone #, email, etc.?
  - Age and gender of the target (if included)
  - Categorization of the victim
  - Categorization of the motive of attacker

# Collection Statistics

Study Period	Summer 2016	Winter 2016-17	Combined
Text Files Recorded	484,185	1,253,702	1,737,887
Classified as Dox	2,976	2,554	5,530
Doxes w/o Duplicates	2,326	2,202	4,528
Manually Labeled	270	194	464

# Outline

- Problem area
- Measurement methodology
- Results and findings
- Discussion and conclusions



# Outline

- Results and findings
  - Doxing targets
  - Doxing perpetrators
  - Effects on social networks

# Doxing Targets

# Victim Demographics

- Taken from the 464 manually labeled doxes
- Only based on data in doxes
- Careful to avoid further harm (e.g. not taking demographic data from OSN accounts)

<b>Min Age</b>	10 years old
<b>Max Age</b>	74 years old
<b>Mean Age</b>	21.7 years old
<b>Gender, Female</b>	16.3%
<b>Gender, Male</b>	82.2%
<b>Gender, Other</b>	0.4%
<b>Located in USA</b>	64.5% (of 300 files that included address)

# Types of Data in Doxes

## Frequently Occurring Data

Category	# of Doxes	% of Doxes*
Address	422	90.1%
Phone #	284	61.2%
Family Info	235	50.6%
Email	249	53.7%
Zip Code	227	48.9%
Date of Birth	155	33.4%

## Highly Sensitive Data

Category	# of Doxes	% of Doxes*
School	48	10.3%
ISP	100	21.6%
Passwords	40	8.6%
Criminal Record	6	1.3%
CCN	20	4.3%
SSN	10	2.6%

\*All numbers from 464 manually labeled doxes

# Doxing Victims by Community

- Categorization of victim based on listed OSN accounts
- 16.2% of victims categorizable into 3 categories

Category	Criteria	# of Labeled	% of Labeled
Hacker	2 or more OSN accounts on hacking sites (e.g. <a href="http://hackforums.net">hackforums.net</a> )	17	3.7%
Gamer	2 or more OSN accounts on gaming sites (e.g. <a href="http://twitch.tv">twitch.tv</a> , <a href="http://minecraftforum.net">minecraftforum.net</a> )	53	11.4%
Celebrity	Labelers recognized target independent of doxing (e.g. Donald Trump, Hillary Clinton)	5	1.1%
Total		75	16.2%

# Doxing Perpetrators

# Doxer Motivations

- Categorization of doxers based on "why I did it" suffixes
- 28.4% of dox motivations categorizable into 4 categories

Category	Criteria	# of Labeled	% of Labeled
Competitive	Demonstrating attacker's capabilities / victim's weaknesses	7	1.5%
Revenge	Because of doxee's actions against doxer (e.g. "you cheated in counterstrike.")	52	11.2%
Justice	Because of doxee's actions against third party (e.g. "you ripped off my friend")	68	14.7%
Political	Because of larger political goal (attacking KKK members or child pornographers)	5	1.1%
Total		132	28.4%

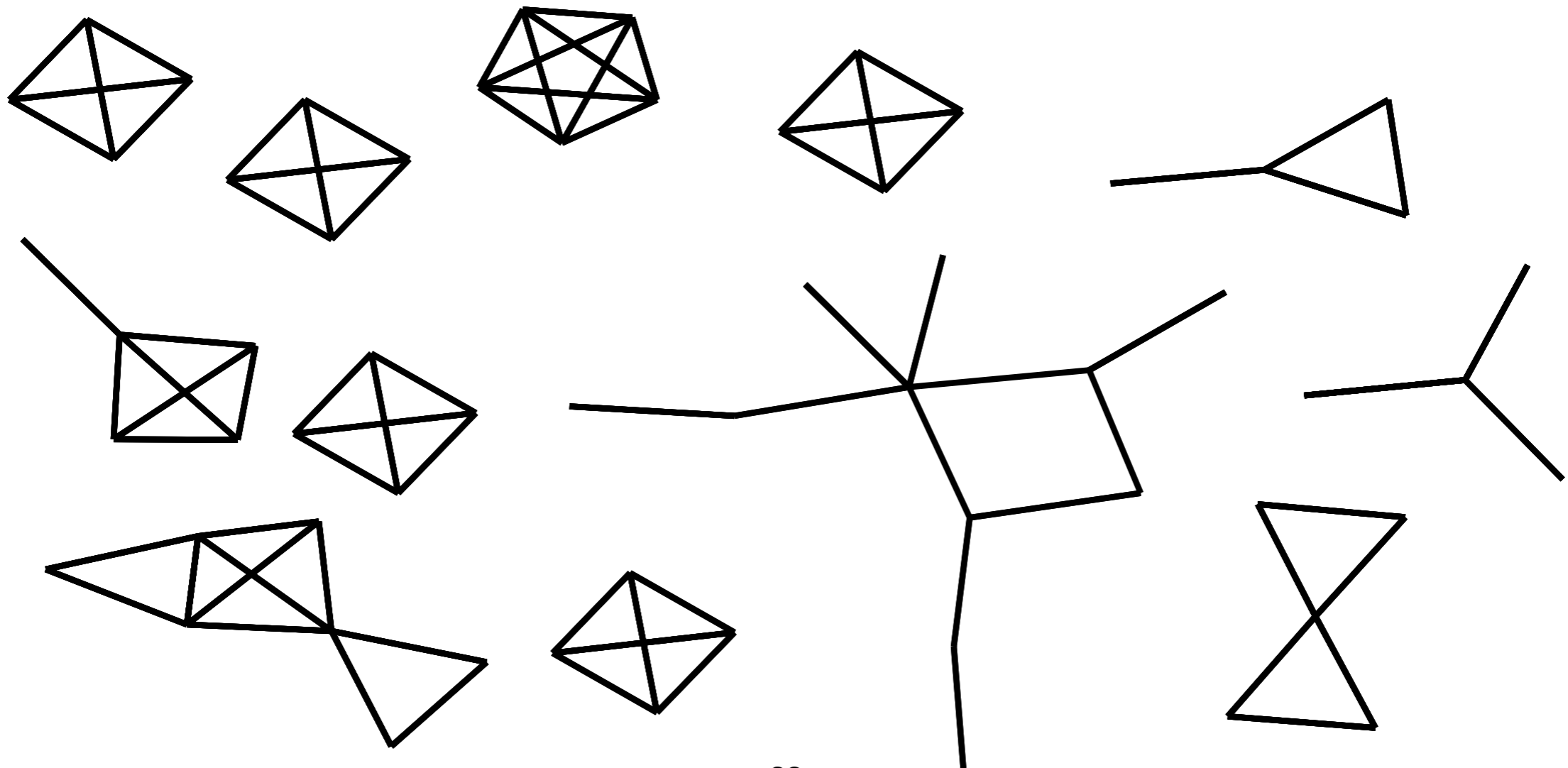
# Doxer Networks

- Looked for doxer networks based on "credit lines"
- ex: "by Alice and @Bob, thx to Charlie (@Charlie for SSN)"
- 251 aliases given, 213 twitter handles
- Undirected graph from doxes and twitter network



# Doxer Networks

- 61 (of 251) aliases appear in cliques of 4 or more
- 34 Twitter accounts were private



# Harms from Doxing

# Effects on OSN Accounts

1. Are OSN accounts in dox files more likely to increase privacy settings?

- 13,392 "background" vs "doxxed" OSN accounts

2. Does OSN abuse filtering reduce the impact of doxing on OSN accounts?

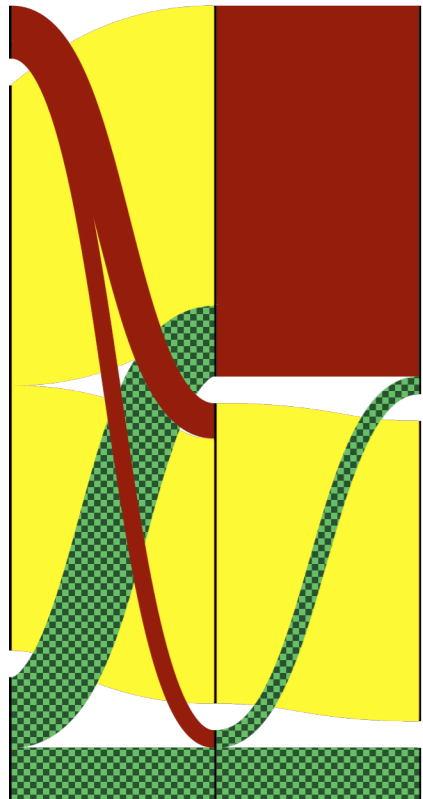
- Before and after increased OSN abuse filtering



# Doxed vs. Non-Doxed Accounts

Account	Condition	% More Private	% More Public	% Any Change	Total #
Instagram	default	0.1	0.1	0.2	13,392
Instagram	doxed, pre-filtering	17.2	8.1	32.2	87
Instagram	doxed, post-filtering	5.7	1.4	9.9	141
Facebook	doxed, pre-filtering	22.0	2.0	24.6	191
Facebook	doxed, post-filtering	3.0	<0.1	3.3	361

# Facebook Statues after Doxing



active

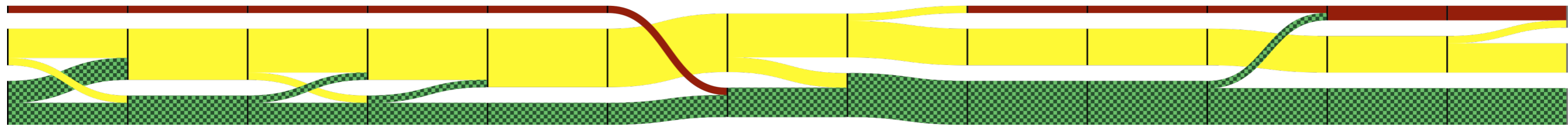


private

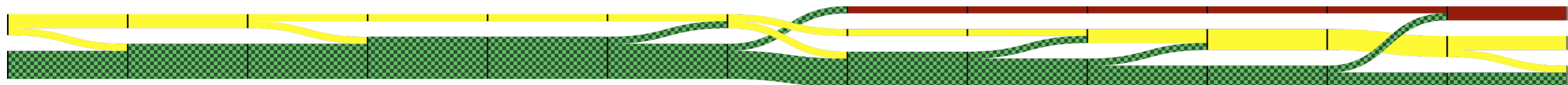


inactive




# Instagram Statues after Doxing



**Instagram accounts that changed status, Pre-filtering (13.8%)**



**Instagram accounts that changed status, Post-filtering (5.0%)**

 active  private  inactive

# Outline

- Problem area
- Measurement methodology
- Results and findings
- Discussion and conclusions

# Using Data to Help Victims

- **Notification of doxing victims**  
"Have I Been Pwned" style service
- **OSN Account protection**  
Notify social networks of doxing, for defenses
- **Anti-SWAT-ing List**  
Additional information for law enforcement to evaluate
- **Anti-Abuse Policies From Dox Distributing Sites**  
Working with Pastebin to increase automated takedowns



# Take Aways

- Automatic dox measurement and classification pipeline
- 1.7m text files, 4,328 doxes, manual labeling of 464
- First quantitative analysis of frequency, targets and contents of doxing online
- Measurement of harm of doxing, via OSN account change