

A Light in the Dark Web

Linking Dark Web Aliases to Real Internet Identities

Ehsan Arabnezhad, Massimo La Morgia, Alessandro Mei, Eugenio Nerio Nemmi and Julinda Stefa

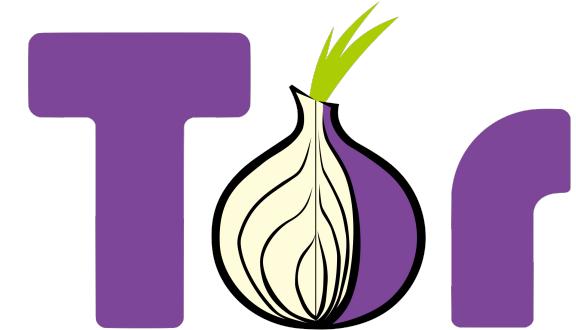
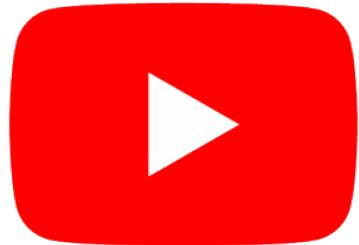


SAPIENZA
UNIVERSITÀ DI ROMA

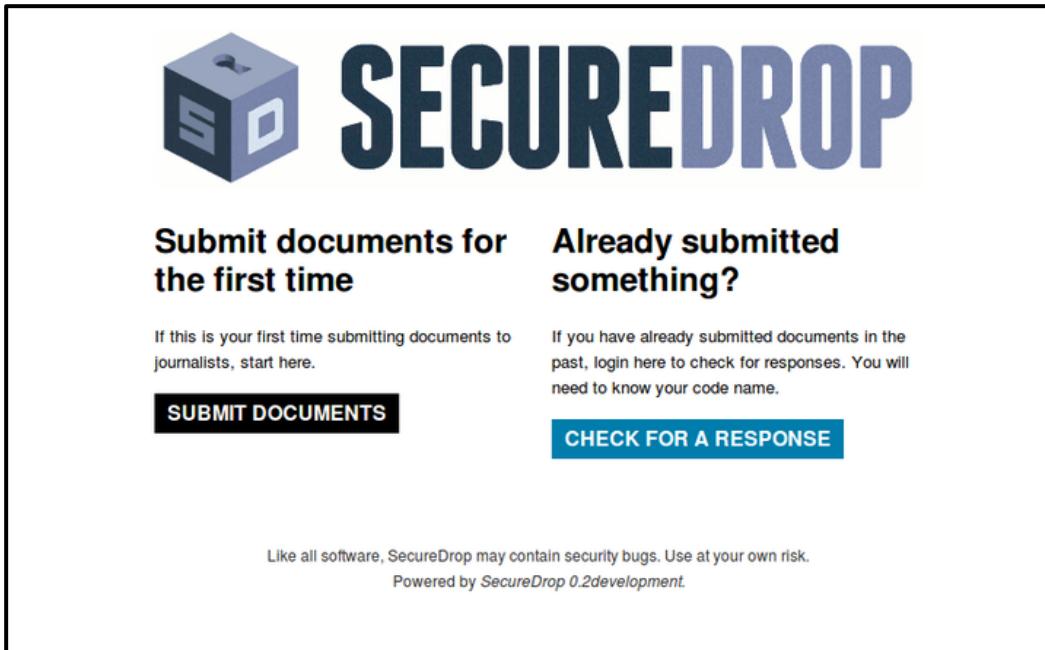
40th IEEE International Conference on Distributed Computing Systems
November 29 - December 1, 2020, Singapore



Social Network



TOR



The SecureDrop homepage features a large logo on the left with a blue cube icon containing a white 'SD' and the word 'SECUREDROP' in bold, blue, sans-serif capital letters. To the right, there are two main sections: 'Submit documents for the first time' with a 'SUBMIT DOCUMENTS' button, and 'Already submitted something?' with a 'CHECK FOR A RESPONSE' button. Below these are two smaller text blocks: one for first-time users and one for returning users. At the bottom, a note about security bugs and a powered-by line are visible.

Submit documents for the first time

If this is your first time submitting documents to journalists, start here.

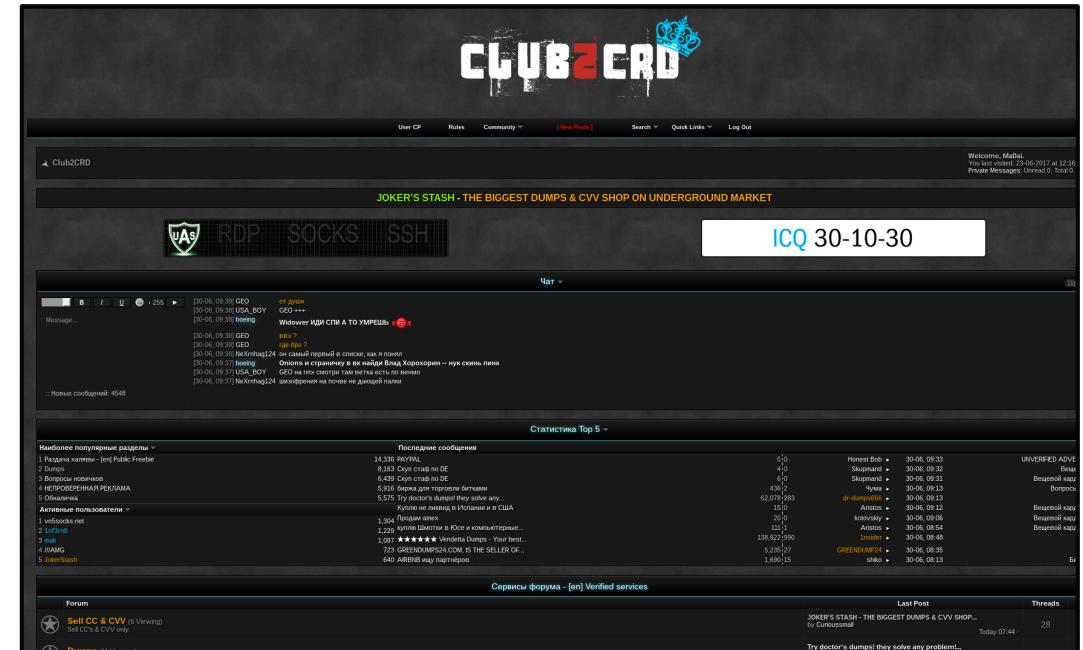
SUBMIT DOCUMENTS

Already submitted something?

If you have already submitted documents in the past, login here to check for responses. You will need to know your code name.

CHECK FOR A RESPONSE

Like all software, SecureDrop may contain security bugs. Use at your own risk.
Powered by *SecureDrop 0.2development*.



The Club2CRD forum interface shows a dark-themed dashboard with various links and a search bar. The main area displays a list of posts from a thread titled 'JOKER'S STASH - THE BIGGEST DUMPS & CVV SHOP ON UNDERGROUND MARKET'. The posts include messages like 'Widower яди СПИ А ТО УМРЕШЬ', 'ICQ 30-10-30', and 'NeXtH4g124'. Below the posts, there are sections for 'Статистика Top 5' (Statistics Top 5) showing user activity and a 'Сервисы форума' (Forum Services) section.

User CP **Rules** **Community** **New Posts** **Search** **Quick Links** **Log Out**

Welcome, Maha! You last visited: 25-06-2017 at 12:18 Private Messages: Unread 0 Total 0

JOKER'S STASH - THE BIGGEST DUMPS & CVV SHOP ON UNDERGROUND MARKET

ICQ 30-10-30

Chat

UAS RDP SOCKS SSH

Статистика Top 5

Последние сообщения

Последнее сообщение	Автор	Время
14.236 [REDACTED]	Honest Bob	30-06-08:33
8,185 Сути страйф DE	Skupmand	30-06-08:32
6,439 Сути страйф DE	Skupmand	30-06-09:31
5,936 бирка для паспорта биометри	lym	30-06-09:13
5,176 бирка для паспорта биометри	dr_diamond	30-06-09:13

Активные пользователи

Активные пользователи	Последний визит
1 vitoocks.net	15:00
2 LetFind	20:00
3 i333	1:15
4 IAMG	1:15
5 JokerStash	1:15

Сервисы форума [en] Verified services

Last Post **Threads**

JOKER'S STASH - THE BIGGEST DUMPS & CVV SHOP... by curiosmail Today 07:44 · 28 Try doctor's dumpst they solve any problem!

- No disclosure of personal information such as gender, nationality, religion.
- English is the only language allowed.

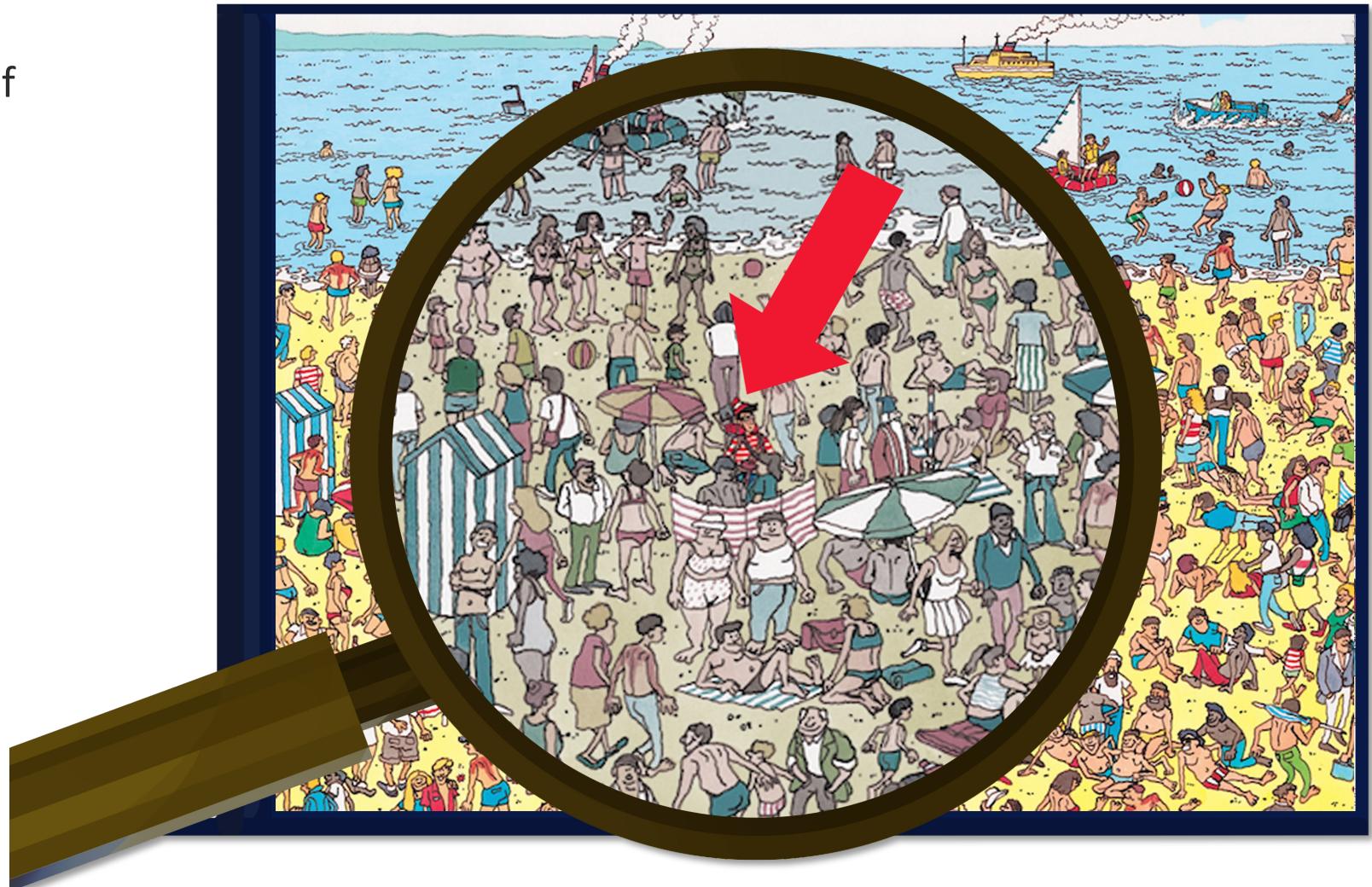
Research Goals

- New methodology to perform Authorship Attribution in the wild:
 - Lots of candidates.
 - Lots of users.
- Methodology applied on the Dark Web.
- Perform an attack on a real case scenario.

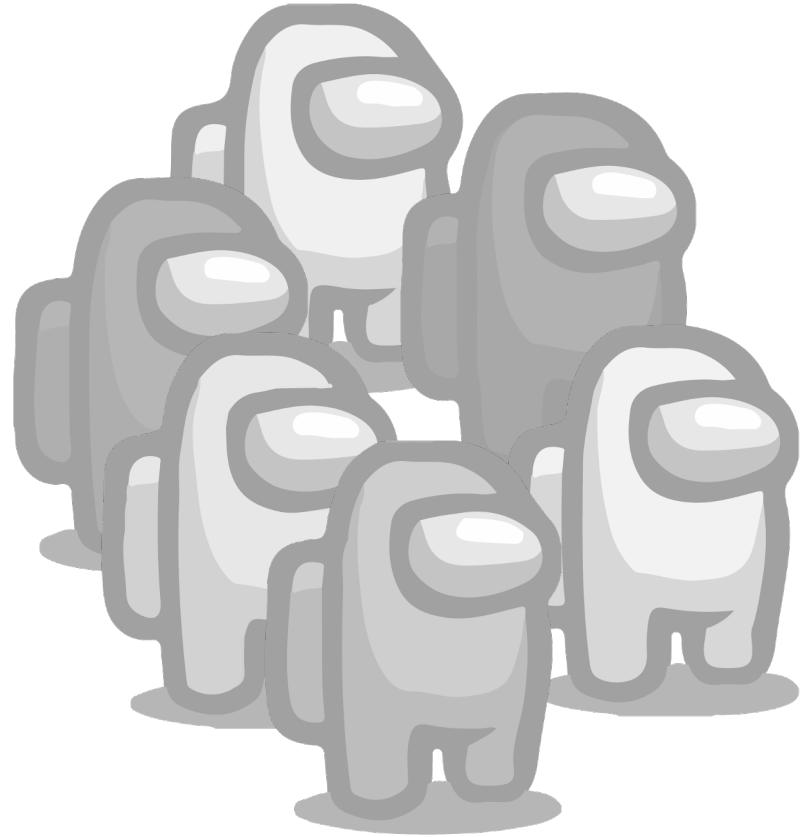
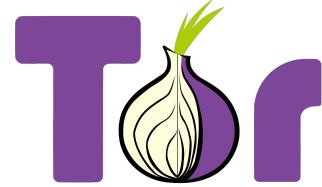


Authorship Attribution

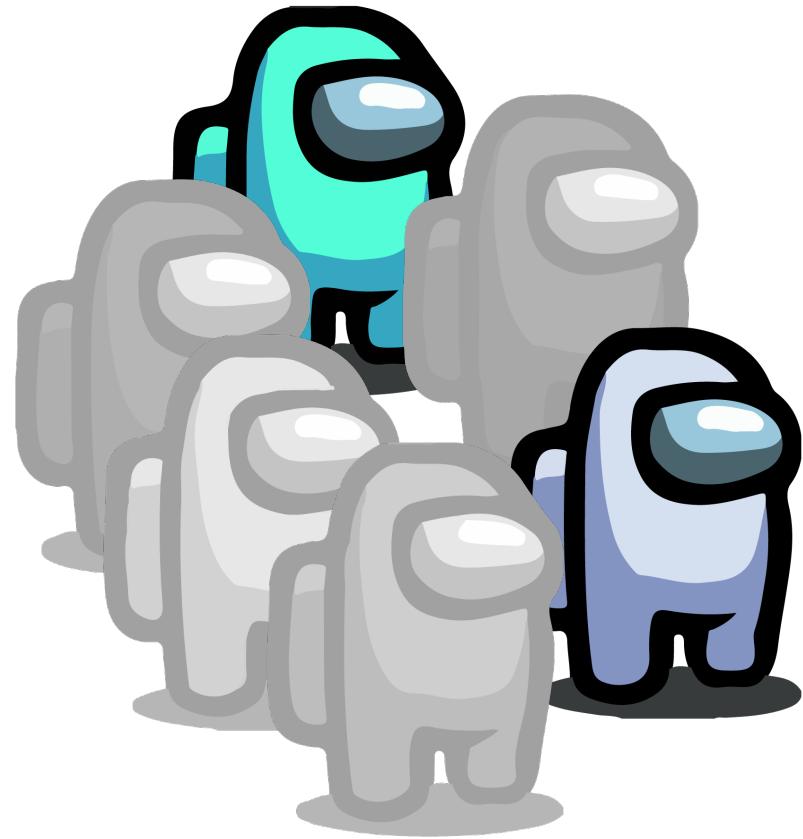
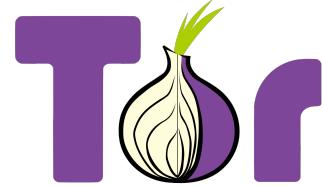
- **Authorship attribution** is the task of identifying the **author** of a given text.



Link users across different platforms



Link users across different platforms





r/DarkNetMarkets



[Anybody bought lsd tabs from Tom and Jerry is it a reliable source](#)

submitted 2 hours ago by

[5 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[Is £160 a good price for a sheet of lsd 150ug sounds a bit cheap?](#)

submitted 22 minutes ago by

[8 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[Is there anyway to access the dark web on an iphone? \[HELP!\]](#)

submitted an hour ago by

[6 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[Monopoly](#)

submitted an hour ago by

[1 comment](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[So this is what everyone's been talking about it 😊](#)

submitted 18 hours ago by

[6 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[Got some from ggpx](#)

submitted 2 hours ago by

[2 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)



[Is cakewallet fast enough for WHM 3 hour time period? \[GUIDE\]](#)

submitted 4 hours ago by

[2 comments](#) [share](#) [save](#) [hide](#) [give award](#) [report](#) [crosspost](#)





r/DarkNetMarkets

- We select the Top 1000 topics.
- We collect all the users that wrote on the topics.
- We download the last 1000 messages for each user:
 - Text
 - Subreddit
 - Timestamp



Reddit user's social analysis

Topic	subreddit(#)	subscriptions(%)	messages(%)	popular subreddit	messages(#)
Culture	18	4.7%	2%	r/science	17,442
Cryptocurrencies	39	3.2%	6%	r/bitcoin	96,407
Drugs	117	15.6%	33.7%	r/DarkNetMarkets	670,483
Entertainment	166	39.1%	22.4%	r/pics	75,454
Financial	15	1.6%	0.9%	r/personalfinance	11,590
Lifestyle/Sports	72	9.9%	9.5%	r/LifeProTips	12,109
News	18	4.8%	4.5%	r/worldnews	89,189
Places	43	1.4%	3%	r/canada	11,291
Politics	24	4%	5.9%	r/politics	119,238
R18+	12	1.6%	4.5%	r/sex	10,676
Psychological help	11	1.7%	0.5%	r/GetMotivated	3,733
Tech/Tor	52	5.4%	3.6%	r/technology	26,919
Videogame	61	7.0%	7.3%	r/gaming	41,183

Targeted Dark Web Forums



A screenshot of the DreamMarket Forum homepage. The header has a blue background with the text "DreamMarket Forum" and "DreamMarket user forum.". Below the header is a navigation bar with links for Index, Search, Register, and Login. A message "You are not logged in." is displayed. On the right, it says "Topics: Acth...". The main content area is titled "Announcements" and shows a table with one row. The table has columns for "Forum", "Topics", "Posts", and "Last post". The single announcement is titled "Announcements" and includes the text "Announcements, updates and news regarding Dream Market". It shows 12 topics, 1,352 posts, and was last updated "Yesterday 18:26:42" by "Gowron".

- The Majestic Garden is a forum where people can share experiences related to the assumption of psychedelic drugs.
- Each vendor has his own thread, where users review the quality of the goods and the vendor.
- Most discussions are related to the assumption of drugs, but there are also sections dedicated to the literature of psychedelic and spiritual experiences, and how-tos on drug cooking.

- The Dream Market was one of the most popular marketplaces on the Dark Web since 2013.
- There are four main sections: Products and Vendor Reviews, Marketplace discussions, Advertising and Promotions, and Scams.
- Although on the Dream Market customers could buy several kinds of stuff, in the forum, most of the messages are about drugs.

Targeted Dark Web Forums

DreamMarket Forum
DreamMarket user forum.

[Index](#) [Search](#) [Register](#) [Login](#)

You are not logged in.

Announcements

Forum	Topics	Posts	Last post
Announcements Announcements, updates and news regarding Dream Market	22	1,412	Today 17:23:38 by EvilMadScientist

Marketplace

Forum	Topics	Posts	Last post
Product and Vendor Reviews Reviews for all vendors and listings	4,752	33,678	Today 19:34:03 by catdog69
MarketPlace Discussion General Discussions	4,329	18,339	Today 20:11:03 by prestigewc
Advertising and Promotions Product advertisements and promotions	4,972	13,457	Today 19:40:27 by shroomzilla
Off Topic A place for discussions about darknet and clearnet	1,117	4,000	Today 17:41:52 by SrgtLulz

Scams

Forum	Topics	Posts	Last post
Scams	1,022	0,316	Today 19:48:13

Dream Market
ichudifyeqm4ldjj.onion
jd6yhuvicivehvt4.onion
t3e6ly3uoif4zcw2.onion
7ep7ackunkzdcw3l.onion

Established 2013

Welcome to
The Majestic Garden

visit  Show unread posts since last visit  Show new replies to your posts.

Profile My Messages Logout

Unread Posts 

3504 Posts 159 Topics [Last post by Malcolm-X in Re: I Will take down TMG... on Today at 07:23:38 am](#)

1762 Posts 171 Topics [Last post by savannacat in Re: I have CGMC Invites ... on Today at 12:04:53 pm](#)

13359 Posts 1132 Topics [Last post by HIGHness in Re: Enough LSD To Supply... on Today at 12:40:56 pm](#)

5949 Posts 362 Topics [Last post by bombheadie in Re: Favorite CURRENT Jam... on Today at 03:39:27 am](#)

2017 Posts 150 Topics [Last post by deedrah12 in Re: Best way to take Mes... on October 23, 2016, 04:39:16 am](#)

Unread Posts 

Safety & Awareness Questions and answers on the safe use of products

Avengers Information, Comments & Discussions

- We download all the messages written on the public area.
- We also collect the associated timestamp.

Data polishing & Text Pre-Processing

Data Polishing

- We remove all the duplicates of the messages.
- We remove emojis.
- We remove quotes from messages.
- We remove 'Edit by username'.
- We remove words that are longer than 34 characters.
- We delete PGP Keys from messages.
- We Normalize URLs.

Pre-Processing

Tokenization

Tokenization is the process of breaking up a stream of text into linguistic units such as words, punctuation, or other meaningful elements.

Lemmatization

Reducing an inflected word to its lemmas (e.g., *am*, *are*, *is* → *be*). Thanks to this standardization, we can analyze words with different inflections as a single item.

Features Extraction

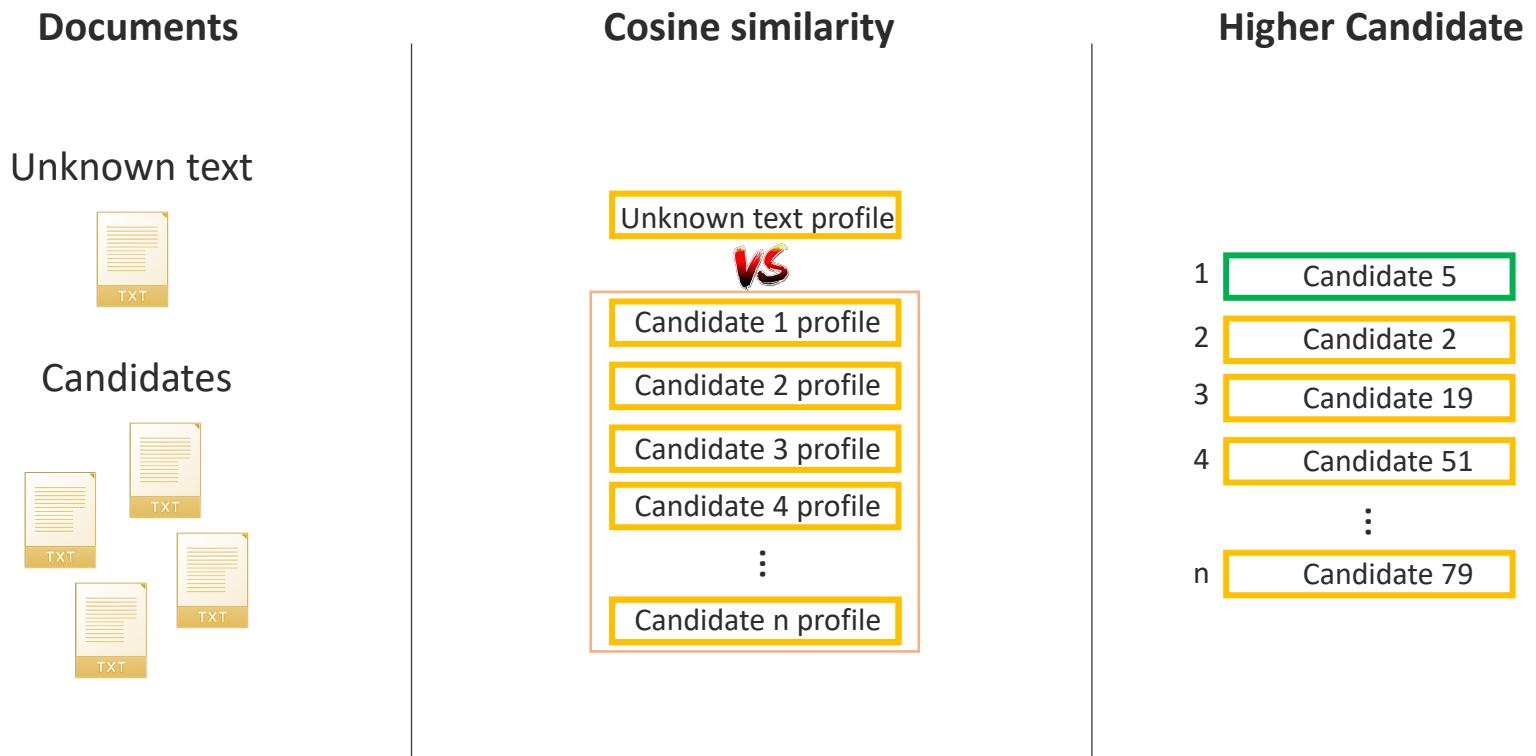
Frequency

- Frequency of punctuation, numbers and special characters.
-

N-Grams

- Char-gram
 - A sub-sequence of N contiguous characters of a larger sequence.
- Word-gram
 - A sequence of N contiguous words in a text.

Methodology

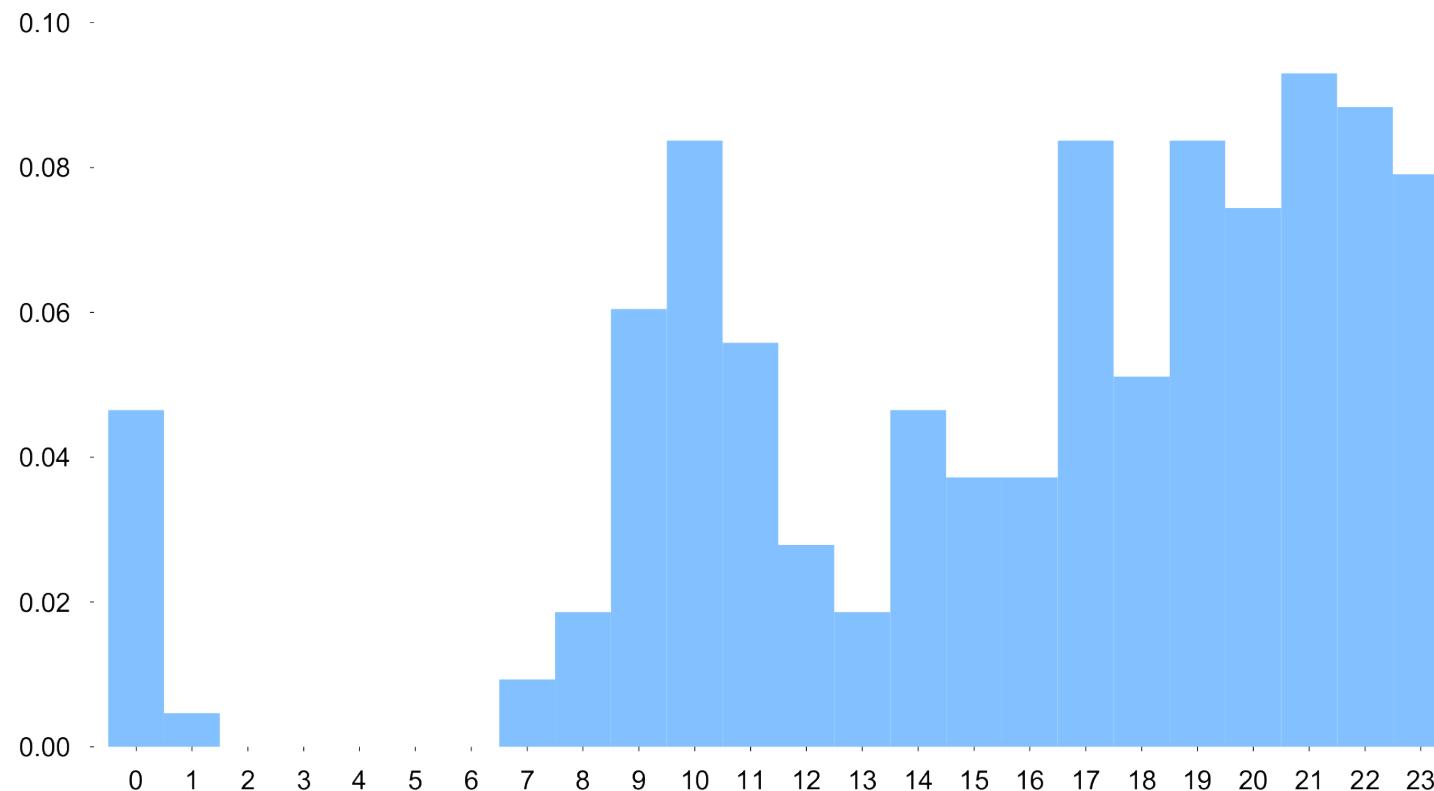


Words vs Accuracy

- Increasing the number of words also increases the accuracy.
- But it needs more messages!

(#)words	K=1(text)
400	16.4%
600	32.5%
800	49.7%
1000	64.6%
1100	68.3%
1200	73.7%
1300	78.6%
1400	81.3%
1500	84.8%
1600	85.3%
1700	87%

User post-time fingerprint



$$P_u = \{P_u[h] | h \in \{0, \dots, 23\}, P_u[h] = \frac{\sum_d a_d(h)}{\sum_{d,h} a_d(h)}\}$$

Comparison

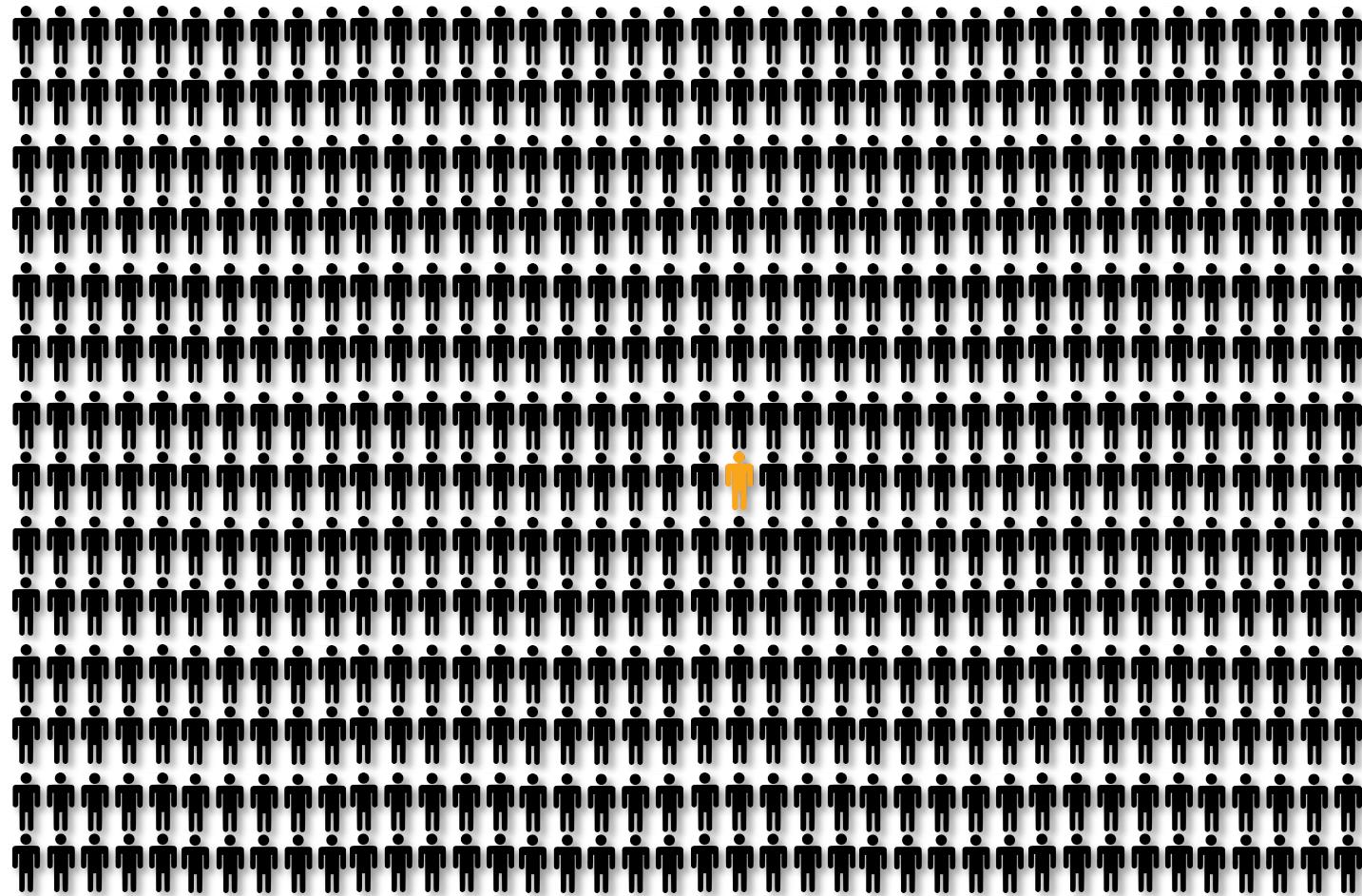
- Using the temporal features allow us to increase the accuracy.
- More than 85% of accuracy with 1500 words.

(#)words	K=1(text)	K=1 (all)
400	16.4%	20%
600	32.5%	37.8%
800	49.7%	55.8%
1000	64.6%	69.6%
1100	68.3%	73.2%
1200	73.7%	76%
1300	78.6%	82.3%
1400	81.3%	84.4%
1500	84.8%	87.7%
1600	85.3%	87.9%
1700	87%	90%

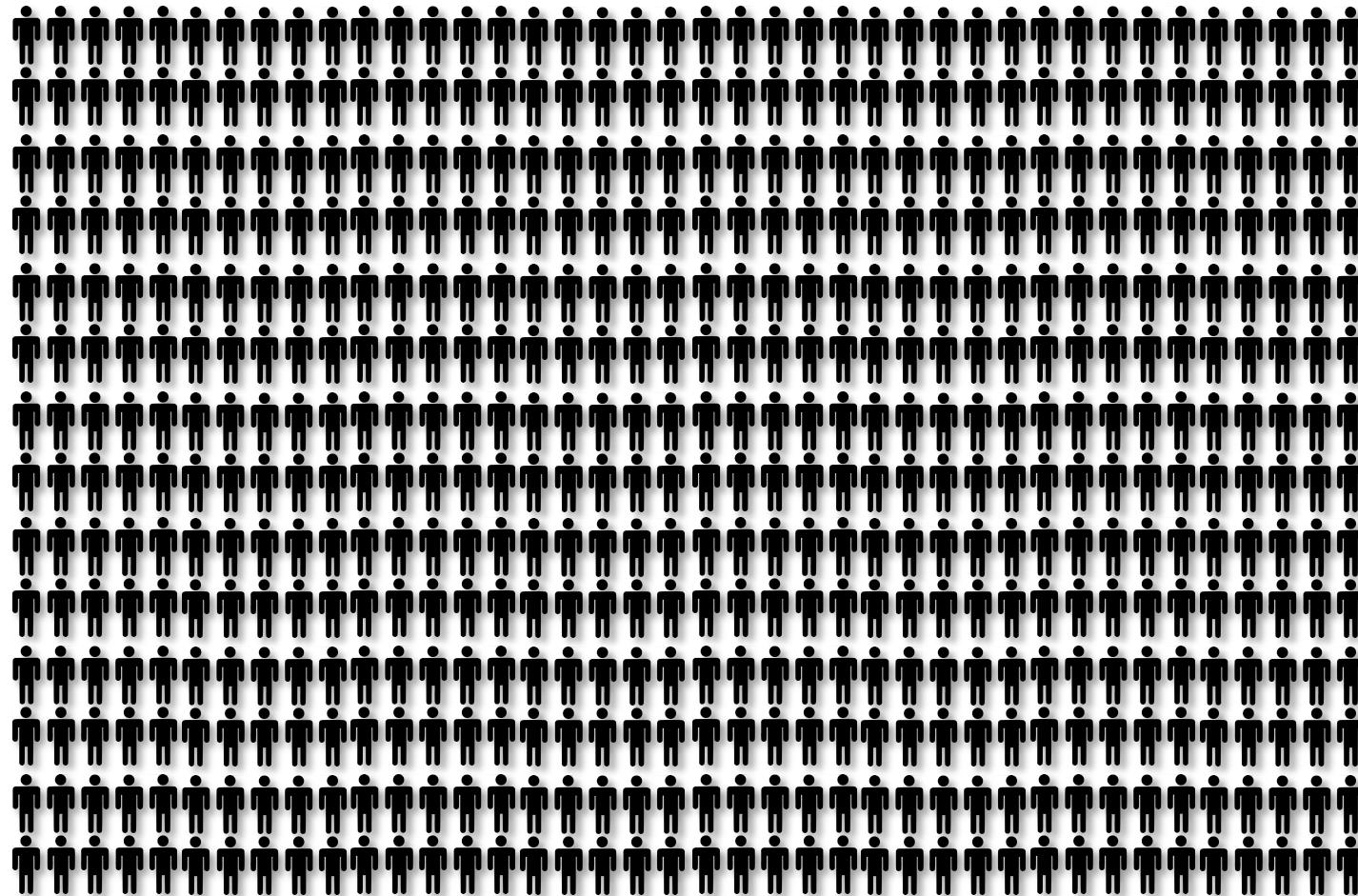


Not Enough!

Too many candidates!



What if the author is not a candidate?



K-best authors!

- We take the first 10 authors!
- With 1500 words, accuracy over 95%.

(#)words	K=1(text)	K=1 (all)	K=10(text)	K=10 (all)
400	16.4%	20%	29.6%	35.5%
600	32.5%	37.8%	51.7%	58.2%
800	49.7%	55.8%	70%	75.2%
1000	64.6%	69.6%	79.7%	84.4%
1100	68.3%	73.2%	83.7%	87.6%
1200	73.7%	76%	87.2%	89.2%
1300	78.6%	82.3%	89.1%	92%
1400	81.3%	84.4%	89.7%	93.4%
1500	84.8%	87.7%	93.4%	95.5%
1600	85.3%	87.9%	94.7%	96.5%
1700	87%	90%	95.7%	97%



Dataset – final numbers

Reddit

Original Users
16,567

Cleaned Users
11, 679

Alter_Ego Users
10,133

DreamMarket Forum

Original Users
6,348

Cleaned Users
178

Alter_Ego Users
66

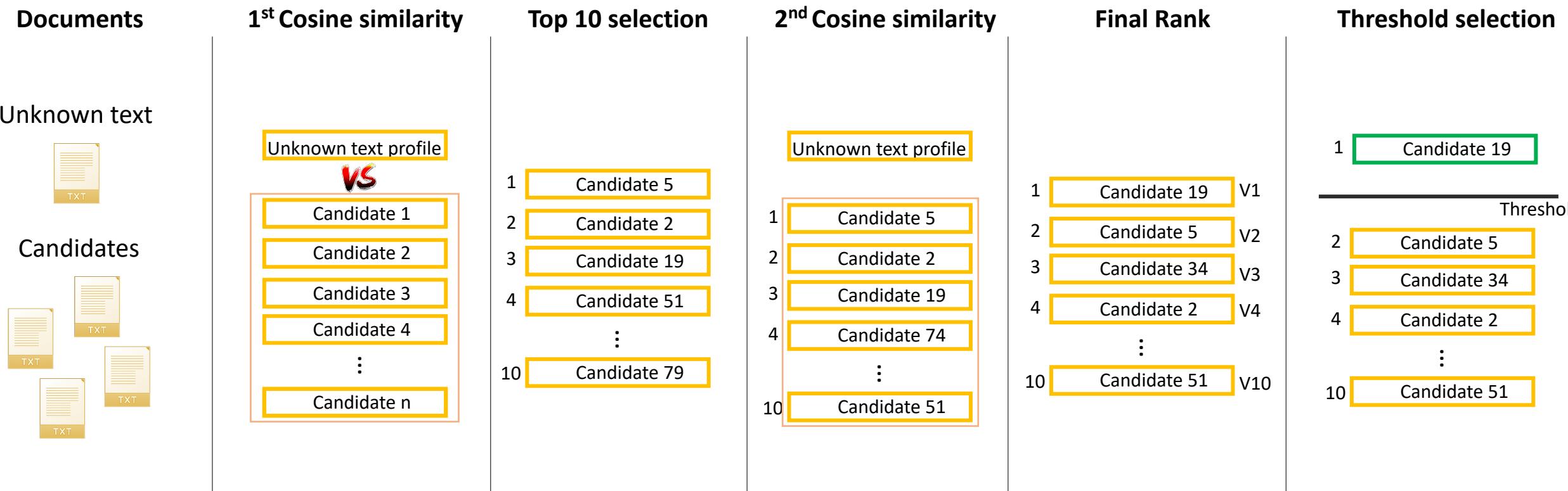
TheMajestic Garden

Original Users
4,709

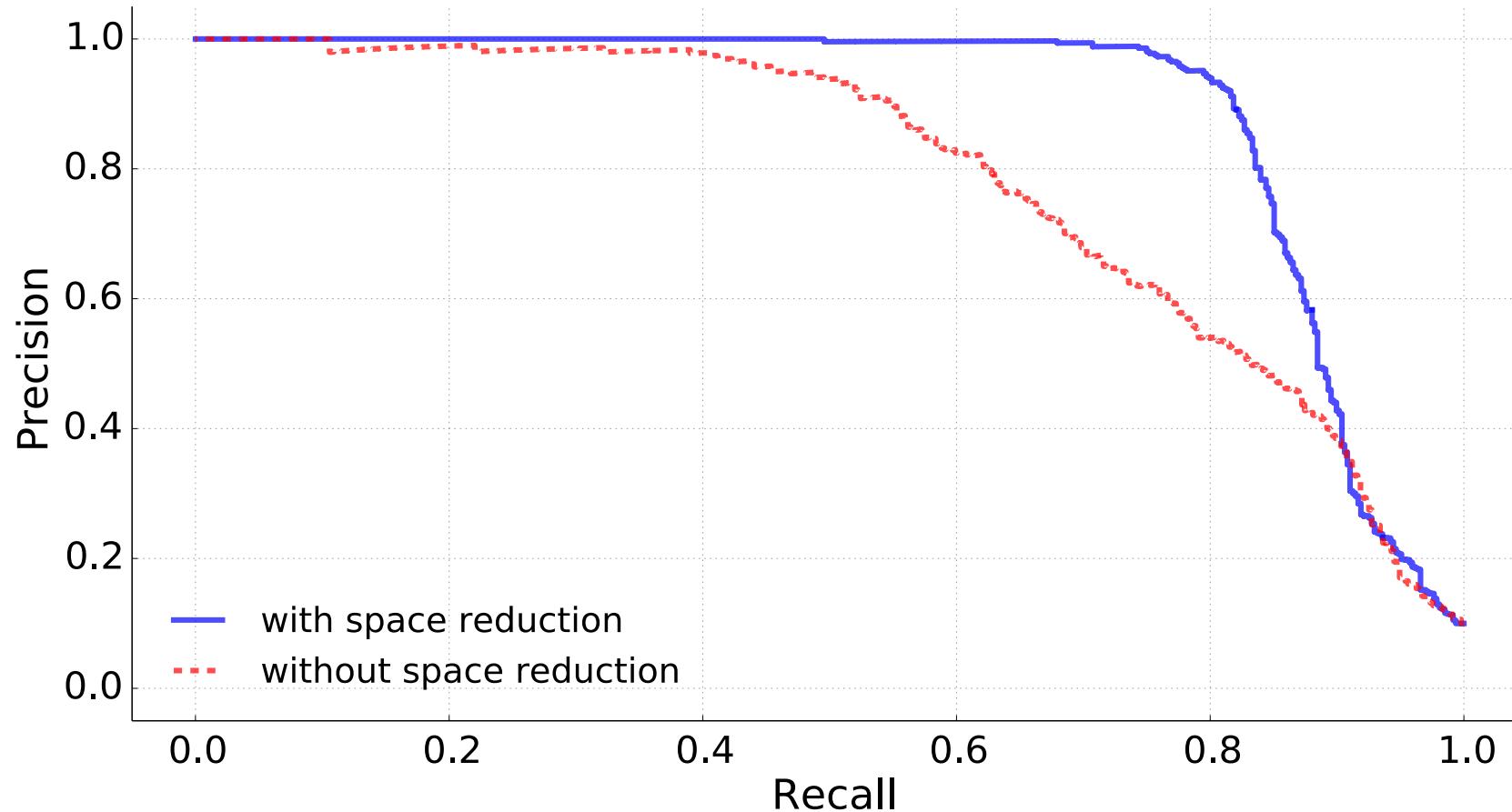
Cleaned Users
422

Alter_Ego Users
196

New Methodology

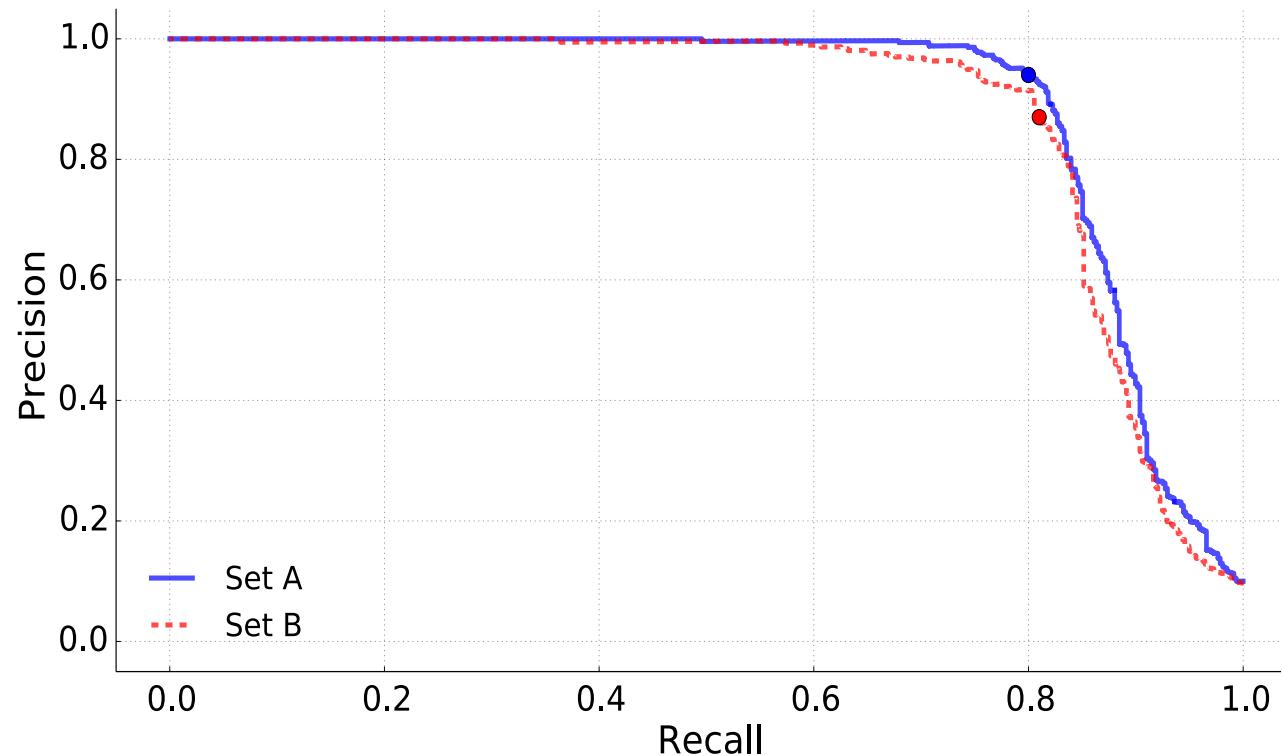


Original AUC vs Space-Reduction AUC



Finding a threshold

- We randomly take 1000 users from AE_reddit.
- We split them into 2 sets of 500 users:
 - Set A
 - Set B
- We determine a threshold on Set A and check if the threshold also holds for Set B.
- **Same threshold gives similar results on both the sets!**



Threshold comparison

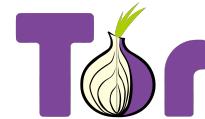
	Forum	threshold	Precision	Recall
• Precision and threshold associated with 80% Recall.	Reddit_A	0.4190	94%	80%
	Reddit_B	0.4210	91%	80%
	DM	0.4096	96%	80%
	TMG	0.4222	94%	80%
• Precision and Recall with the selected threshold.	Reddit_A	0.4190	94%	80%
	Reddit_B	0.4190	87%	82%
	DM	0.4190	98%	78%
	TMG	0.4190	90%	84%

Results in the wild – True pairs



Reddit users

[...] Yes, I'm X on DM
(Dream Market, Ed.), I
don't have MDMA now
but... []



Dark Web users

[...] But no MDMA man,
for that you have to
wait... []

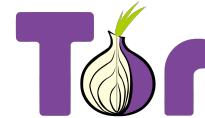
18 True pairs

Results in the wild – Probably True pairs



Reddit users

[...] I bought it from X, I'm sure he ships in Canada because I live there... []



Dark Web users

[...] If you are in Canada, you can try X, is one of the best vendor. Never had a problem! [...]

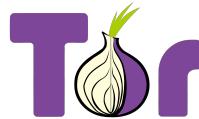
2 Probably True pairs

Results in the wild – Unclear



Reddit users

[...] I don't do Heroin,
and as for now, I never
thought about trying it
... []



Dark Web users

[...] I don't know about
his Heroin, never
bought it. I only smoke
weeds, can't help you
... []

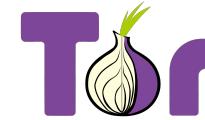
20 Unclear

Results in the wild – False



Reddit users

[...] Raise and born here
in Poland ...[]



Dark Web users

[...] I'm an American
citizen...[]

7 False

Let me present you... John Doe!

- He is 27 years old.
- He lives in Edmonton, Canada.
- He lives with his parents.
- ..And he has brother!
- He lost his job because of drugs abuse.
- He is in a 2 years relationship.
- He loves to play online video games.
- He uses a Samsung Galaxy S4.
- Finally... **we even know where he goes to drink.**



Thank you!



SAPIENZA
UNIVERSITÀ DI ROMA