HOLISTICA Vol 12, Issue 1, 2021, pp.27-40 DOI:10.2478/hjbpa-2021-0003

ILLICIT ACTIVITIES BENEATH THE SURFACE WEB: INVESTIGATING DOMESTIC EXTREMISM ON ANONYMOUS SOCIAL MEDIA PLATFORMS

Maurice DAWSON[1]*
Andreas VASSILAKOS[2]
Jose Luis CASTANON REMY[3]
Tenace Kwaku SETOR[4]

Received: February 2021 | Accepted: March 2021 | Published: April 2021
Please cite this paper as: Dawson, M., Vassilakos, A., Castanon Remy, J.L. Setor, T.K. (2021) Illicit activities beneath the surface web investigating domestic extremism on anonymous social media platforms, *Holistica Journal of Business and Public Administration, Vol.12, Iss.1, pp.27-40*

Abstract

At the beginning of 2021, the Internet was used to spread words to incite insurrection and violence through social media, incited a pro-Trump mob riot into a U.S. Capitol building. Furthermore, due to recent acts of domestic terrorism in Texas, New Zealand, and California, police authorities have begun investigating social media presence that premeditated harmful acts. In all three instances, the shooters posted their manifesto online and had a presence on 8chan. The common thread among all shooters is their identification as a white nationalist and their social media site affiliation with Infinitechan, where they hid their radical ideas. Similarities in the shooters' profiles include their perceived viewpoints of population groups regarding their political, ethnic, and social identities. This paper will provide insight into the forums where domestic terrorists spread their agendas. It will also set the foundation for further research towards a strategic algorithm that compiles and analyses relevant users' profiles by using OSINT and data analytics techniques).

Keywords: Extremism; Open Source Intelligence; Homeland Security; Counterterrorism; Crime; Cyber; Cyber Security; Software Development

[1] *University Fernando Pessoa, Science and Technology Faculty, Porto, Portugal.*
* *Corresponding author.*
[2] *Illinois Institute of Technology, Center for Cyber Security and Forensics Education, Chicago, IL, USA.*
[3] *Technical University of Madrid, Escuela Técnica Superior de Ingenieros Informáticos, Madrid, Spain.*
[4] *University of Nebraska Omaha, College of Information Science & Technology, Omaha, NE, USA.*

1. Introduction

Reddit is a network of online communities based on people's interest. Figure 1 displays this website that is broken up into more than a million communities known as subreddits that /r/. The subreddits that show begin to display racial discrimination, American nationalism, and other forms of bigotry can be found with /r/politicallyincorrect/. This subreddit is the politically incorrect section with 5.6k members. Researchers at the Pew Research Centre's Internet & American Life Project state that 6% of alone adult users are Reddit users (Duggan & Smith, 2013.

Figure 1 Reddit screenshot

Source: https://www.reddit.com/r/politicallyincorrect/

One of the prominent online platforms is 4chan, which is displayed in Figure 2. It is a forum that hosts over seven million users, along with a representation of their activities outside this electronic medium (Bernstein, Monroy-Hernandez, Harry, Andre, Panovich & Vargas, 2011). Activities showcased in 4chan include demonstrations of hacktivism (like actions of the "Anonymous" group), distributed denial of service attacks to shut down companies like Mastercard and PayPal in support of Wikileaks, and protest against groups like the Church of Scientology (Bernstein, Monroy-Hernandez, Harry, Andre, Panovich & Vargas, 2011).

Figure 2 4chan screenshot

Source: https://www.4chan.org

4chan has many threads; however, the Politically Incorrect thread is where hate speech can be found directed at different groups such as ethnic minorities. Figure 3 shows an overview of this thread; however, upon further inspection politically motivated messages with the swastika, the flag of the Confederate States of America (CSA), and other symbols that represent oppression for various groups are widely used.

Figure 3 4chan Politically Incorrect thread screenshot

Source: https://www.4chan.org

On September 4th, 2019, James Watkins, the owner of 8chan appeared before Congress, the Committee on Homeland Security, and the U.S. House of Representatives in Washington, DC to provide insight on the platform following the incidents of domestic terrorism during 2019. Within the report mentioned above, it is stated that "Why extremism thrives in the current American political landscape remains unknown and deeply problematic. The celebration and exaggeration of victimhood exacerbates social divides and amped-up fear destroys rational debate. 8chan has no silver bullet, no magic cure to end extremism in America today. It knows that censorship and suppression do little to cure underlying societal ills. Thus, it remains committed to the promise of the First Amendment, to allow for unfettered free discussion, and to work with law enforcement when tragedies occur" (Committee on Homeland Security 8chan Inquiry, 2019). The platform, according to the report, currently is not continuing its activities. Before, it was hosting ideas of radical forum members and extremist propaganda.

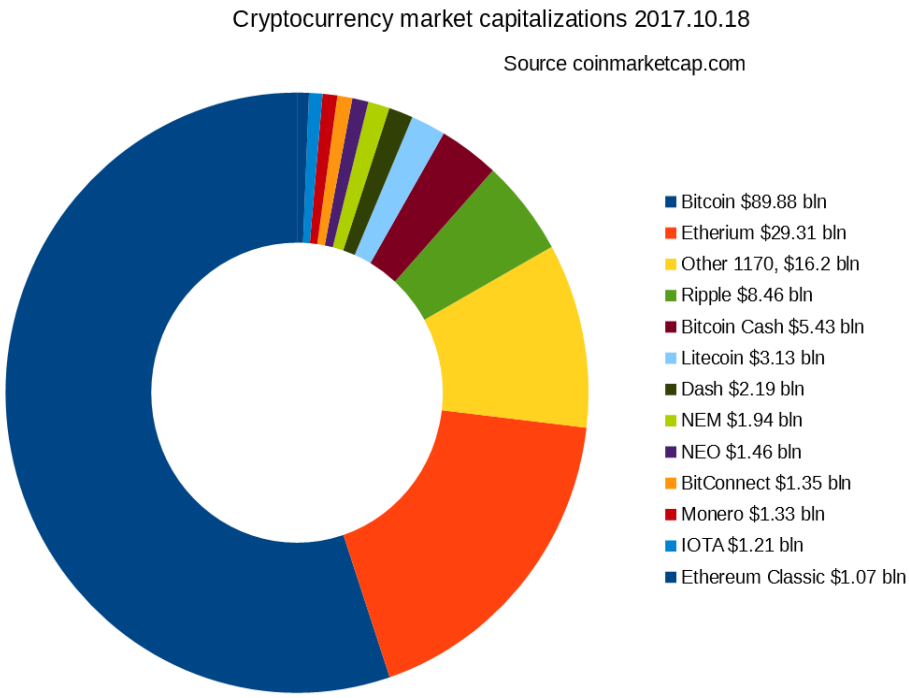
With the ever-growing domestic terrorist events occurring that challenge the fabric of safety in American society, appropriate tools must be used to counter nefarious activities (Dawson, Kisku, Gupta, and Sing, 2016). As manifestos grow on these anonymous message boards, policing is essential to understanding possible domestic attacks. Hate speech and radical ideas are extensively being spread behind layers of secrecy; advancement in the control of such communication means is vital.

2. Digital currency

The currency has been an integral part of everyday life for centuries. People have moved from the barter system of trade to gold and valuable materials, to modern day metallic, paper, and bank money. Money functions as a medium of exchange covering the insufficiencies of the barter system, as a unit of account measuring the value of services and goods, and as a value indicator enabling its reuse and redistribution among people (Shoaib, 2013).

Digital Currency can be defined as a type of digital money, which is issued and controlled by its developers and is an acceptable medium of exchange among members of specific virtual groups. There are several different types of digital currency (see Figure 4).

Figure 4 Cryptocurrency Market Capitalizations

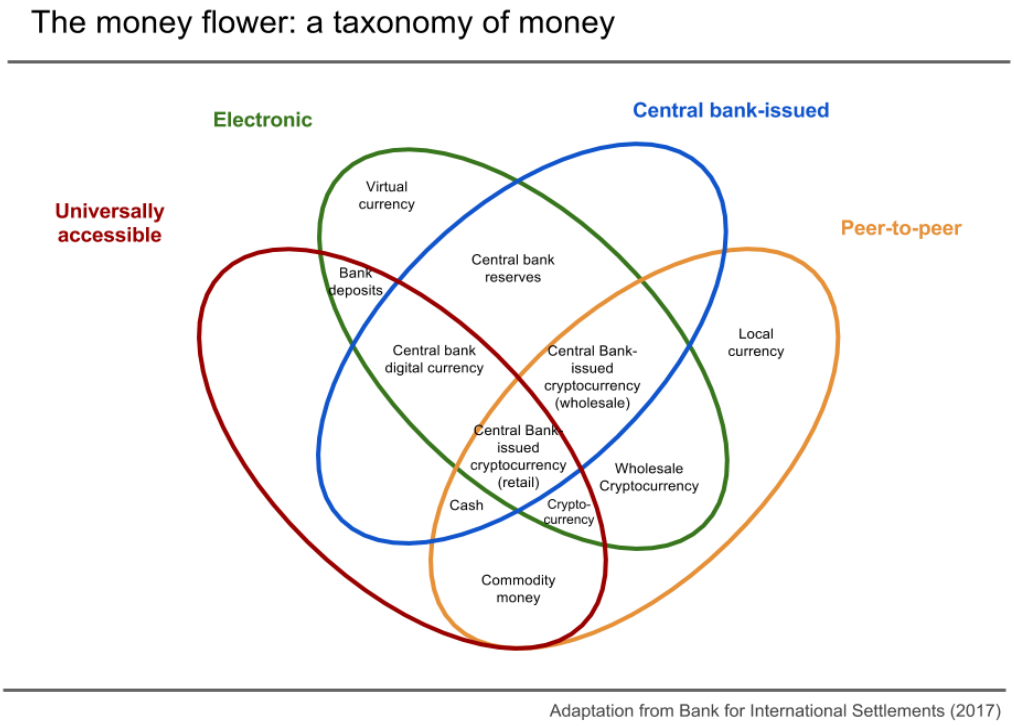


Source: https://commons.wikimedia.org/wiki/File:Cryptocurrency-market-capitalizations-2017-10-18.png. Licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication license

Flaws also characterize digital currency, but of a different nature to the ones of traditional currency. They are not stable, and they could potentially impact the reputation of its users and banks negatively when they are associated with it and are not issued or controlled by a central source (Shoaib, 2013).

Paper money is seemingly one of the most popular and widely accepted currency types, but it has some flaws such as the inability to be anonymous. Additionally, it is susceptible to theft, and counterfeit currency is continuously an issue the government has to face. Digital currency is trying to provide a solution for these inadequacies and understanding the flow of money is essential to uncovering how these currencies interact within different spheres (Shoaib, 2013).

Figure 5 Money Flower



Source: https://commons.wikimedia.org/wiki/File:Money_flower.png. Licensed under a Creative Commons Attribution-Share Alike 4.0 International license

Digital currency's existence, despite the conceptualization of its original form, has been tied to illicit activities in the Deep/Dark Web. Essentially that means that nefarious users have been using this type of money so that they can stay under the radar and hide any trails their actions would potentially leave behind.

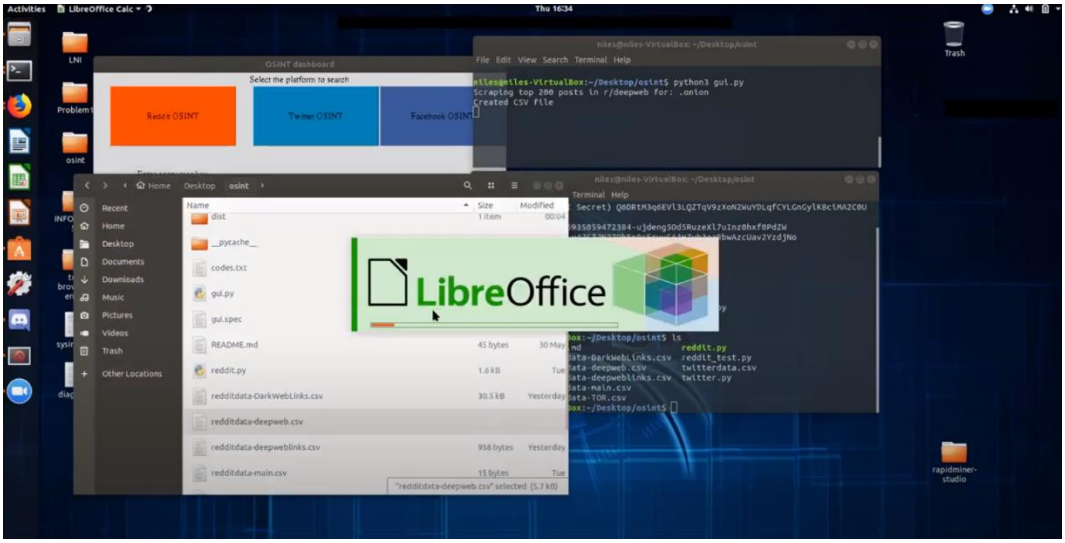
There have been several instances where digital currencies have been used in a nefarious or unethical way. Traffickers have been using the Dark Web to retain anonymity. People were trying to make quick investments creating fictional demand for several types of digital money (Griffith, 2018).

3. OSINT

Open Source Intelligence (OSINT) has been used from the United States (U.S.) Intelligence Community (IC) to state-sponsored groups. OSINT has been coupled with psychological warfare to increase tactics such as population control or influencing specific groups. In the recent election a researcher polled over 9,000 people and they saw a total of 5,000 ads (Lapowsky, 2018). Combined the people saw a combined 5 million paid ads on

Facebook between September 28 and November 8, 2016 (Lapowsky, 2018). Looking at the number of paid ads and Russian created groups such as Black Lives Matter to promote civil unrest.

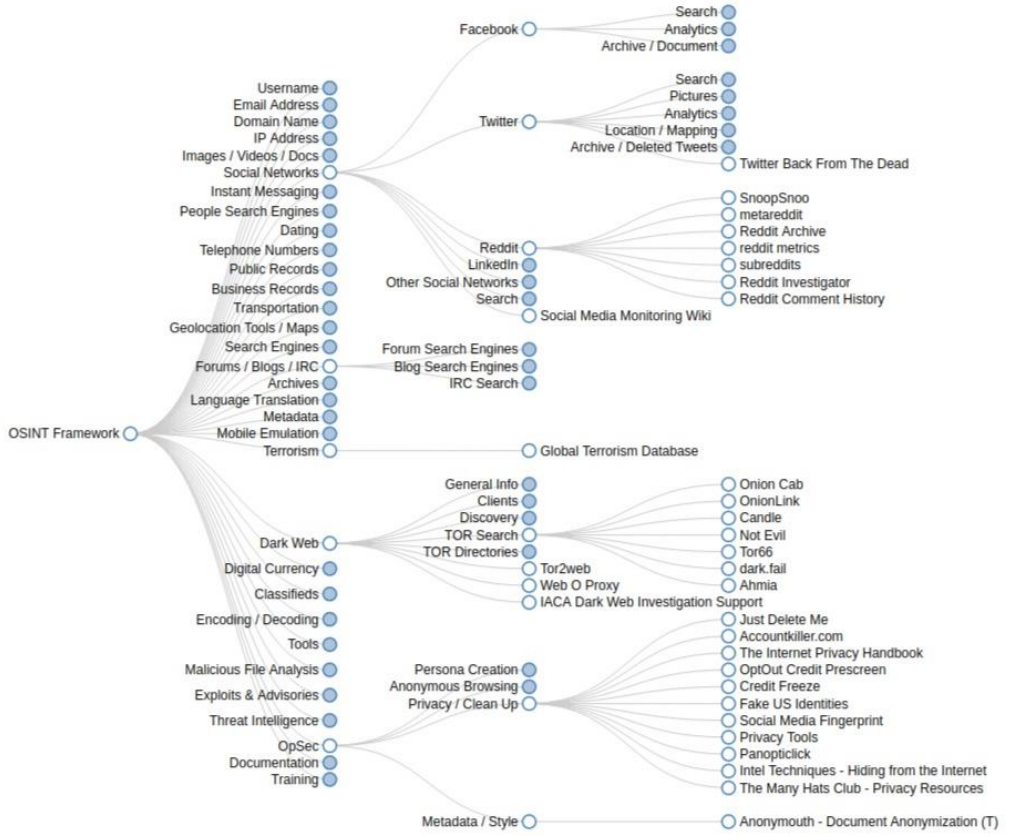
Figure 6 Social Media Social OSINT Application



Source: Authors capture, 2021

In Figure 6 displayed is an application created by University of Missouri - Saint Louis (UMSL) students that show advanced cybersecurity operations and targeted intelligence that meet the needs of the NSA and Department of Homeland Security (DHS) Centre of Academic Excellence (CAE) in Cyber Defence Education (CDE) (Wang, Dawson, and Williams, 2018).

Figure 7 Implementing the OSINT Framework

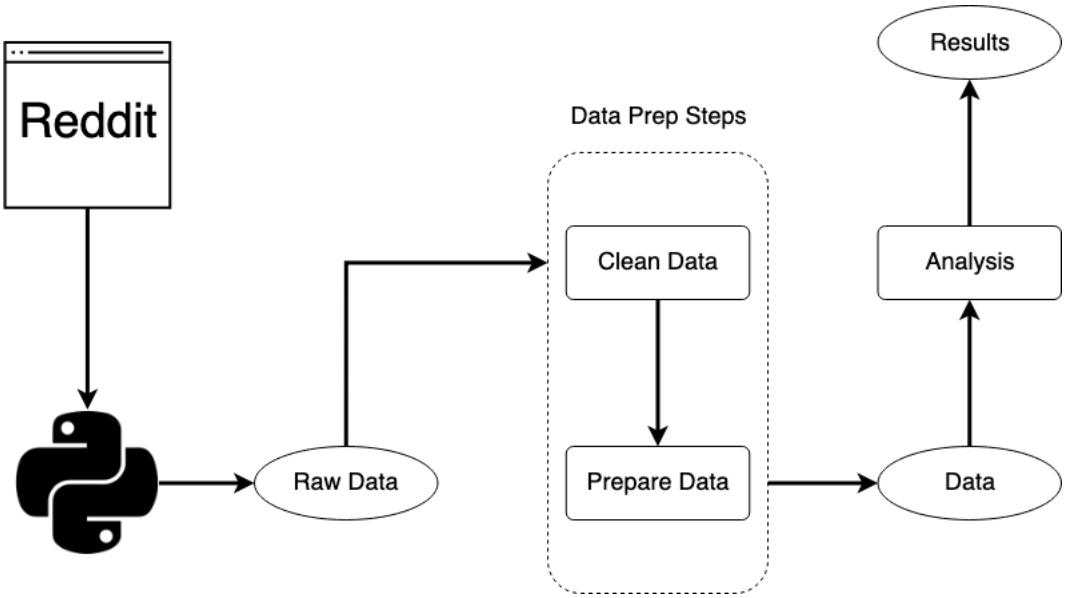


Source: https://osintframework.com/

4. OSINT Project & Algorithm

The script is based on four different libraries. Those libraries are the following, praw, requests, csv, and time. Praw library will let the script use the API of Reddit, based on the credentials of a Reddit user. The following credentials from a Reddit user are needed in order to get the code running fine, username, password, user agent, client id, and client secret. All these parameters can be found under user settings, under privacy and security settings, and inside the link app authorization. This configuration cannot be found unless the user is logged in. Explain requests, csv, and time.

Figure 8 Algorithm flowchart

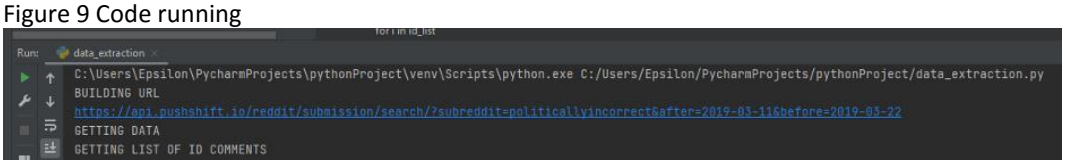


Source: Authors capture, 2021

There are three parts to the script - Fetch Input Values, Fetch Data, and Build CSV Document. The script accepts a subreddit URL for a given date range. The subreddit is a specific community on reddit where posts dedicated to a particular topic, and comments and sub comments associated with the posts are shared by users. As an exemplar, our code fetches sample data associated with the "politically incorrect" subreddit between a given date range. Through the request's library, we retrieve a JSON file with information from all the Reddit posts from "politically incorrect" and between two dates. The JSON file has different keys and their corresponding values including ID. Each post in Reddit has a unique ID to identify the post. With that ID, we can retrieve each post and all related data. The IDs are stored in a list.

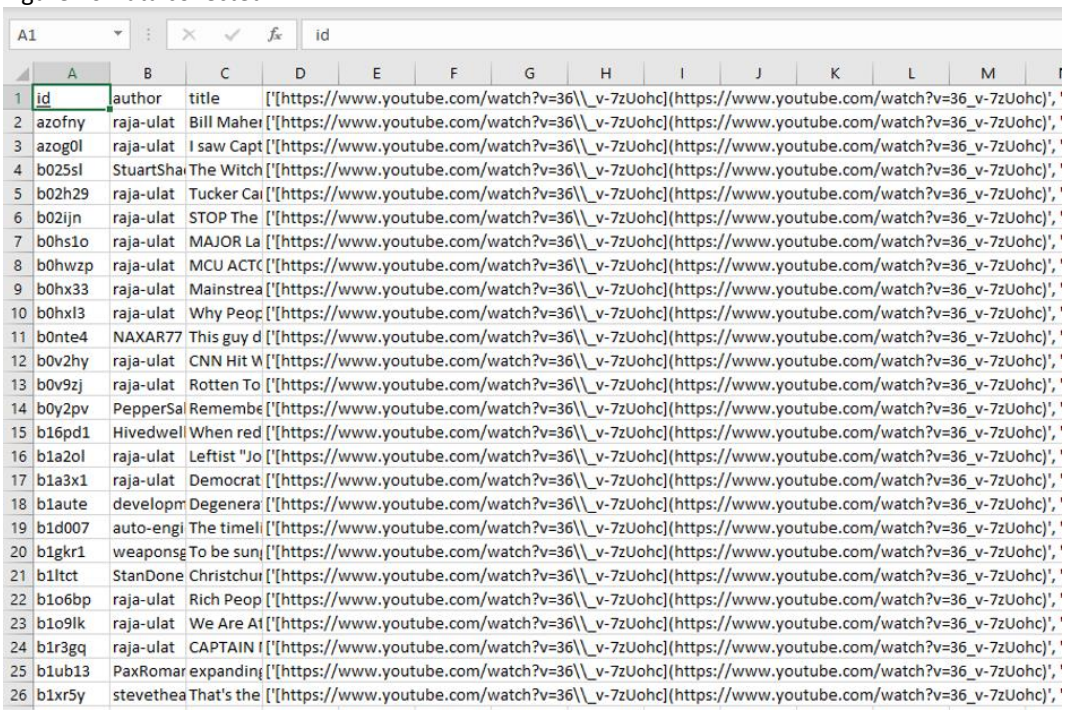
The second part of the code, Fetch Data. retrieves data associated with each post. The script searches for all the comments with the corresponding IDs. The third part of the script Build CSV document, builds a CSV. The CSV is structured in columns namely IDs, Author, Title, and Comments. The code finally builds a CSV file as the output (Figure 10).

Figure 9 Code running



Source: Authors capture, 2021

Figure 10 Data collected



	A	B	C	D
1	id	author	title	['https://www.youtube.com/watch?v=36_v-7zUohc',
2	azofny	raja-ulat	Bill Maher	['https://www.youtube.com/watch?v=36_v-7zUohc',
3	azog0l	raja-ulat	I saw Capt	['https://www.youtube.com/watch?v=36_v-7zUohc',
4	b025sl	StuartSha	The Witch	['https://www.youtube.com/watch?v=36_v-7zUohc',
5	b02h29	raja-ulat	Tucker Ca	['https://www.youtube.com/watch?v=36_v-7zUohc',
6	b02ijn	raja-ulat	STOP The	['https://www.youtube.com/watch?v=36_v-7zUohc',
7	b0hs1o	raja-ulat	MAJOR La	['https://www.youtube.com/watch?v=36_v-7zUohc',
8	b0hwzp	raja-ulat	MCU ACTC	['https://www.youtube.com/watch?v=36_v-7zUohc',
9	b0hx33	raja-ulat	Mainstrea	['https://www.youtube.com/watch?v=36_v-7zUohc',
10	b0hxl3	raja-ulat	Why Peop	['https://www.youtube.com/watch?v=36_v-7zUohc',
11	b0nte4	NAXAR77	This guy d	['https://www.youtube.com/watch?v=36_v-7zUohc',
12	b0v2hy	raja-ulat	CNN Hit W	['https://www.youtube.com/watch?v=36_v-7zUohc',
13	b0v9zj	raja-ulat	Rotten To	['https://www.youtube.com/watch?v=36_v-7zUohc',
14	b0y2pv	PepperSal	Remembe	['https://www.youtube.com/watch?v=36_v-7zUohc',
15	b16pd1	Hivedwell	When red	['https://www.youtube.com/watch?v=36_v-7zUohc',
16	b1a2ol	raja-ulat	Leftist "Jo	['https://www.youtube.com/watch?v=36_v-7zUohc',
17	b1a3x1	raja-ulat	Democrat	['https://www.youtube.com/watch?v=36_v-7zUohc',
18	b1aute	developm	Degenera	['https://www.youtube.com/watch?v=36_v-7zUohc',
19	b1d007	auto-engi	The timeli	['https://www.youtube.com/watch?v=36_v-7zUohc',
20	b1gkr1	weaponsg	To be sun	['https://www.youtube.com/watch?v=36_v-7zUohc',
21	b1ltct	StanDone	Christchur	['https://www.youtube.com/watch?v=36_v-7zUohc',
22	b1o6bp	raja-ulat	Rich Peop	['https://www.youtube.com/watch?v=36_v-7zUohc',
23	b1o9lk	raja-ulat	We Are At	['https://www.youtube.com/watch?v=36_v-7zUohc',
24	b1r3gq	raja-ulat	CAPTAIN I	['https://www.youtube.com/watch?v=36_v-7zUohc',
25	b1ub13	PaxRomar	expandin	['https://www.youtube.com/watch?v=36_v-7zUohc',
26	b1xr5y	stevethea	That's the	['https://www.youtube.com/watch?v=36_v-7zUohc',

Source: Authors capture, 2021

The code the researchers used for this item can be found below. All credentials used by the researchers to operate the API have been replaced with the “NA” value for privacy purposes.

```
import praw
import requests
import csv
import time

'''
Get all posts' ids between both given dates
Given the list of ids, PRAW looks for the comments of each post
Match each id with comments
Add everything to a csv
'''


# GET INPUT VALUES
# 1. SUBREDDIT
# 2. DATE BEFORE
# 3. DATE AFTER

#subreddit = input('Enter subreddit: ')
```

```
#date_after = input('Enter date after: ')
#date_before = input('Enter date before: ')
#print('LOOKING FOR POSTS WITHIN ' + subreddit
#          + 'AFTER ' + date_after + 'AND BEFORE ' + date_before)

subreddit = 'politicallyincorrect'
date_before = '2019-03-22'
date_after = '2019-03-11'

# CREDENTIALS TO USE THE API
reddit = praw.Reddit(client_id='NA',
                     client_secret='NA',
                     user_agent='NA',
                     username='NA',
                     password='NA')

print('BUILDING URL')
# BUILD THE URL
urrl = 'https://api.pushshift.io/reddit/submission/search/?' \
          'subreddit=' + subreddit + '&' \
          'after=' + date_after + '&' \
          'before=' + date_before
print(urrl)

# GET DATA (JSON)
print('GETTING DATA')
json_data = requests.get(urrl).json()


# BUILDING CSV FORMAT
# COLUMNS ARE:
# 1. ID
# 2. AUTHOR
# 3. TITLE
# 4. COMMENTS
comment_list = []
csvData = [['id', 'author', 'title', comment_list]]
id_list = []

print('GETTING LIST OF ID COMMENTS')
for i in json_data['data']:
          id_list.append(i['id']) # CREATING A LIST OF ID COMMENTS
```

```
id_comment_dic = {}

t_b = time.time()
#print('GETTING COMMENTS FROM EACH POST %s' % t_b)
print('GETTING COMMENTS FROM EACH POST')
for i in id_list:
	submission = reddit.submission(id=i) # GET SUBMISSIONS FROM REDDIT POST ID GIVEN
	submission.comments.replace_more(limit=0)
	for comment in submission.comments.list():
	comment_list.append(comment.body) # ADD EACH COMMENT TO THE COMMENT LIST
	id_comment_dic [i] = comment_list # MATCH EACH LIST OF COMMENTS WITH ITS ID(POST - COMMENTS)
t_e = time.time()
#print('END GETTING COMMENTS %s' % t_e)
print('END GETTING COMMENTS')
print('TOTAL TIME TAKEN %s SECS' % (t_e - t_b))

# BUILDING DATA
print('BUILDING DATA')
for i in json_data['data']:
	csvData.append([i['id'], i['author'], i['title'], id_comment_dic.get(i['id'])]) # APPEND ALL DATA TO A CSV FILE

print('BUILDING DOCUMENT')
# BUILDING CSV DOCUMENT
with open('data.csv', 'w') as csvFile:
	writer = csv.writer(csvFile)
	writer.writerows(csvData) # WRITING ROWS TO THE CSV FILE
csvFile.close() # CLOSE FILE
```

5. Future directions

A feature that needs to be done is to clean and prepare data attached to the CSV. This can be done either by the same script or with another script. Doing things with another script suppose a problem of concurrency. The first script should be finished when the second one has started. Steps were taken to prepare, and clean data should be done based on a preliminary analysis of the data stored in the CSV. Once the problem related to data cleaning and preparation is solved, we can focus on the next problem, the algorithm of selection. This is the most important step to focus on. We need an algorithm to identify and select relevant information inside the prepared data. We have different ideas that might be interesting for this specific research.

Basically, it is possible to analyse data obtained in different ways. Simple features we think might need to be done in order to improve analysis are word count, word analysis, and meaning. This procedure may lead to different methods like word lists, where a list of words can be used to find different tendencies. Based on meaning ideas, these lists can be filled with results from a sentence and word analysis.

From previous ideas, it is possible to develop a method to create profiles, so it is possible to follow the activity of different users. These profiles can be built up based on different patterns found in data retrieved from these users. There are also the possibilities to add machine learning techniques and build a model able to build profiles automatically. We think profiles are the most interesting feature we can add. These profiles can be lately correlated to the people behind these accounts.

References

Balduzzi, M., & Ciancaglini, V. n.d.. Cybercrime in the Deep Web. In Black Hat EU, Amsterdam 2015. Retrieved from https://www.blackhat.com/docs/eu-15/materials/eu-15-Balduzzi-Cybercrmine-In-The-Deep-Web-wp.pdf.

Bergman, M. K. (2001). White paper: the deep web: surfacing hidden value. *Journal of electronic publishing*, 7(1).

Bernstein, M. S., Monroy-Hernandez, A., Harry, D., Andre, P., Panovich, K., & Vargas, G. n.d.. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community. 8.

Best, C. (2011). Challenges in open source intelligence. *Intelligence and Security Informatics Conference (EISIC), IEEE*: 58-62).

Cardenas-Haro, J. A., & Dawson, M. (2017). Tails Linux Operating System: The Amnesiac Incognito System in Times of High Surveillance, Its Security Flaws, Limitations, and Strengths in the Fight for Democracy in *Security Solutions for Hyperconnectivity and the Internet of Things*, 260-271. IGI Global.

Committee on Homeland Security 8chan Inquiry before the U.S. House of Representatives, 106th Cong. (2019).

Dawson, M., & Cárdenas-Haro, J. A. (2017). Tails Linux Operating System: Remaining Anonymous with the Assistance of an Incognito System in Times of High Surveillance. *International Journal of Hyperconnectivity and the Internet of Things (IJHIoT)*, 1(1), 47-55.

Dawson, M., Kisku, D. R., Gupta, P., Sing, J. K., & Li, W. (Eds.). (2016). *Developing Next-Generation Countermeasures for Homeland Security Threat Prevention*. IGI Global.

Denis, M., Zena, C., & Hayajneh, T. (2016). Penetration testing: Concepts, attack methods, and defense strategies. *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*. doi:10.1109/lisat.2016.7494156

Duggan, M., & Smith, A. (2013). 6% of online adults are reddit users. *Pew Internet & American Life Project*, 3: 1-10.

Griffith, E. (2018). The Dark Side of Crypto Revolution. Accessed December 10, 2018. https://www.wired.com/story/the-dark-side-of-the-crypto-revolution/.

Harris, S., & Meyers, M. (2002). *CISSP*. McGrawHill/Osborne

Janczewski, L., & Colarik, A. (2007). *Cyber Warfare and Cyber Terrorism. Hershey*. PA: IGI Global. doi:10.4018/978-1-59140-991-5

Lapowsky, I. (2018). How Russian Facebook Ads Divided and Targeted US Voters Before the 2016 Election. Accessed December 3, 2018. https://www.wired.com/story/russian-facebook-ads-targeted-us-voters-before-2016-election/.

Lima, D. (2016). Lessons to be learned from Adult FriendFinder hack of 412M accounts. Accessed December 19, 2018. https://www.bizjournals.com/southflorida/news/2016/11/21/lessons-to-be-learned-from-adult-friendfinder-hack.html.

Nagle, A. (2017). *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.

Osint framework. n.d.. Accessed January 9, 2021. https://osintframework.com/.

Pingle, B., Mairaj, A., & Javaid, A. Y. (2018). Real-World Man-in-the-Middle (MITM) Attack Implementation Using Open Source Tools for Instructional Use. *2018 IEEE International Conference on Electro/Information Technology (EIT)*. doi:10.1109/eit.2018.8500082

Qin, J., Zhou, Y., Lai, G., Reid, E., Sageman, M., & Chen, H. (2005). The dark web portal project: collecting and analysing the presence of terrorist groups on the web, in *Proceedings of the 2005 IEEE international conference on Intelligence and Security Informatics*, 623-624. Springer-Verlag.

Shoaib, M., Ilyas, M., & Khiyal, M. S. (2013). Official digital currency. *Eighth International Conference on Digital Information Management (ICDIM 2013)*. doi:10.1109/icdim.2013.6693982

Telford, T., & Timberg, C. (2018). Marriott discloses massive data breach affecting up to 500 million guests. Accessed December 19, 2018. https://www.washingtonpost.com/business/2018/11/30/marriott-discloses-massive-data-breach-impacting-million-guests/?noredirect=on&utm_term=.717f57eb3cea.

Wang, J., & Yang, T. (2018). Subliminal Channel and Digital Currency Pay Security. *2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. doi:10.1109/icitbs.2018.00041

Wang, P., Dawson, M., & Williams, K. L. (2018). Improving Cyber Defense Education through National Standard Alignment: Case Studies. *International Journal of Hyperconnectivity and the Internet of Things (IJHIoT)*, 2(1), 12-28.