

Old Dominion University

ODU Digital Commons

Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics

NSF Research Experiences for Undergraduates (REU) Programs 2022

Summer 2022

Discovering the Traces of Disinformation on Instagram

Haley Bragg

Christopher Newport University, haley.bragg.19@cnu.edu

Michele C. Weigle (Mentor)

Old Dominion University, mweigle@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/reu2022_computerscience



Part of the [Computer Sciences Commons](#), and the [Social Media Commons](#)

Recommended Citation

Bragg, Haley and Weigle, Michele C. (Mentor), "Discovering the Traces of Disinformation on Instagram" (2022). *Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics*. 5.

https://digitalcommons.odu.edu/reu2022_computerscience/5

This Presentation is brought to you for free and open access by the NSF Research Experiences for Undergraduates (REU) Programs 2022 at ODU Digital Commons. It has been accepted for inclusion in Computer Science: Research Experiences for Undergraduates in Disinformation Detection and Analytics by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.



Discovering the Traces of Disinformation on Instagram

Haley Bragg

Mentor: Dr. Michele Weigle

August 4, 2022



1

Overview

Disinformation on Instagram

• Problem: Understudied Instagram Disinformation Campaigns

○ Senate Intelligence Committee's Report on the Internet Research Agency

- Instagram: 187 million engagements
- Facebook: 77 million
- Twitter: 73 million

https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf



<https://www.counterhate.com/malalgorithm>

<https://www.counterhate.com/disinformationdozen>

- What Makes Studying Instagram So Difficult?

Twitter

Instagram

Posts shared through retweeting and quote tweeting	Posts shared through reposting on Instagram stories
Link sharing	No direct links in Instagram posts
No edit feature (yet)	Captions, tags, and locations may be edited
Easier to archive (backdoor)	Harder to archive (no backdoor)
330 million monthly active users	1.21 billion monthly active users

● What Role Does Web Archiving Play?

- Disinformation Dozen responsible for 65 percent of anti-vaxx content
 - We need a way to preserve this content when accounts get banned

1. Joseph Mercola
2. Robert F. Kennedy, Jr. **BANNED**
3. Ty and Charlene Bollinger **BANNED**
4. Sherri Tenpenny **BANNED**
5. Rizza Islam **BANNED**
6. Rashid Buttar **BANNED**
7. Erin Elizabeth **BANNED**
8. Sayer Ji **BANNED**
9. Kelly Brogan
10. Christiane Northrup **BANNED**
11. Ben Tapper **BANNED**
12. Kevin Jenkins **BANNED**



<https://www.counterhate.com/disinformationdozen>

2

How Well Do Instagram Mementos Replay?

A Disinformation Dozen Example

- Robert F. Kennedy Jr.

- Most mementos of Disinformation Dozen

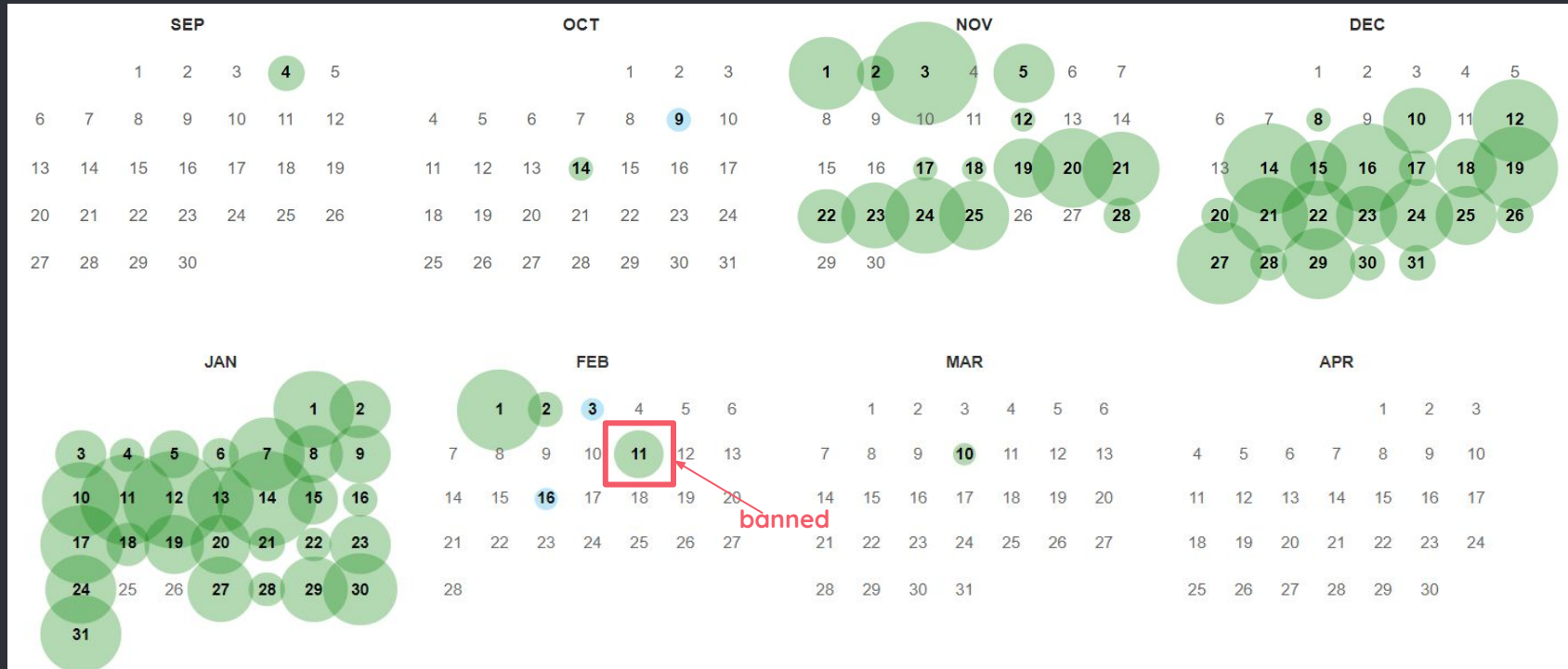
- Instagram as primary platform for spreading disinformation
- Banned February 2021



Mementos of Robert F. Kennedy Jr.'s social media account pages

	Instagram	Facebook	Twitter
Internet Archive	598	60	193
Archive-It	0	0	34
archive.today	0	0	55
Library of Congress	2	0	1
UK Web Archive	0	0	3
Australian Web Archive	0	0	0
Portuguese Web Archive	1	0	0
Total	601	60	286

- Wayback Machine Calendar View 2020-2021



● Redirect to Login Page vs. Success

INTERNET ARCHIVE
WayBackMachine

Loading...

http://www.instagram.com/robertfkennedyjr |
09:02:04 January 01, 2021

Got an HTTP 302 response at crawl time

Redirecting to...

https://www.instagram.com/accounts/login/



Instagram

Phone number, username, or email

Password

Log In

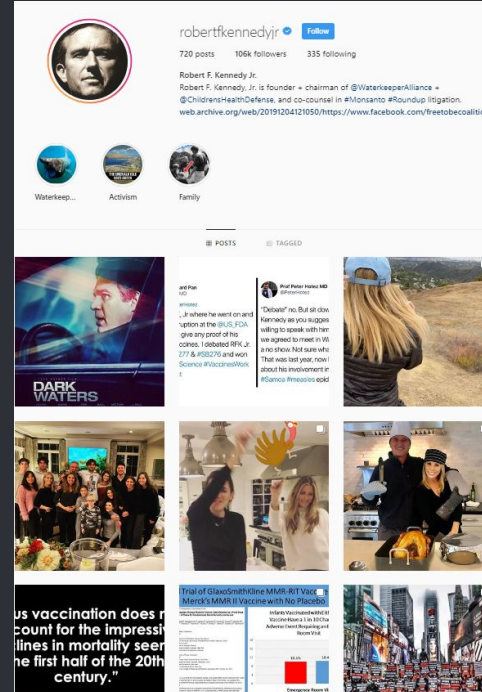
OR

 Log in with Facebook

Forgot password?

Don't have an account? [Sign up](#)

Get the app.



Post
images

“Blue” Memento Count: **24**

• More Redirects than Portrayed

Replayable Memento Count: **6**

	Number of Mementos	Percentage of Total Mementos
Identified Redirects	574	95.99%
Warc/Revisits to Redirects	18	3.01%
Total Redirects	592	99.0%
Identified Successes	6	1.0%
Warc/Revisits to Successes	0	0.0%
Total Successes	6	1.0%
Total Mementos	598	100.0%

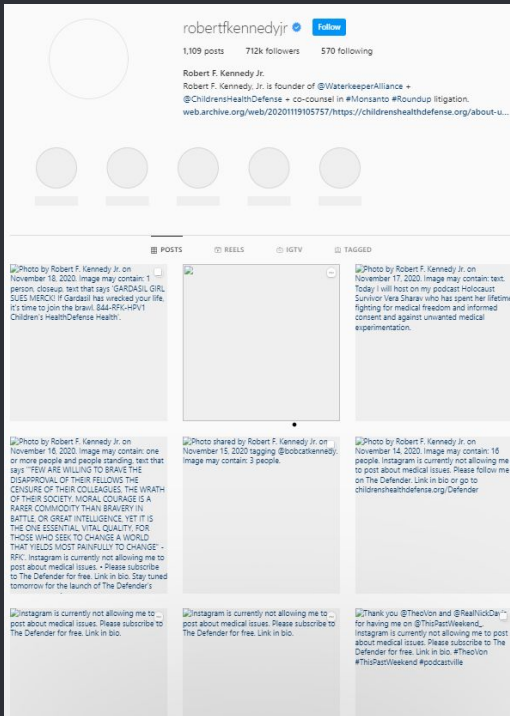
Table 2: Redirects vs. successes for [instagram.com/RobertFKennedyJr](https://www.instagram.com/robertfkennedyjr/)

```
20200417071047 https://www.instagram.com/robertfkennedyjr/ text/html 200 RPLKSY2U2SKARJ3VRQY0YC4KL4V5JVHJ
20200516085443 https://www.instagram.com/robertfkennedyjr/ text/html 200 XXV6JEW3MULPPXFDL7HBG6XWY6WBF5CS
20200731234947 https://www.instagram.com/robertfkennedyjr/ text/html 302 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ
20200824160554 https://www.instagram.com/robertfkennedyjr/ text/html 302 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ
20200825001651 https://www.instagram.com/robertfkennedyjr/ text/html 302 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ
20200904140717 https://www.instagram.com/robertfkennedyjr text/html 301 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ 1
20200904140717 https://www.instagram.com/robertfkennedyjr/ text/html 302 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ
20201009014906 https://www.instagram.com/robertfkennedyjr/ warc/revisit - 3I42H3S6NNFQ2MSVX7XZKYAYSCX5QBYJ
```

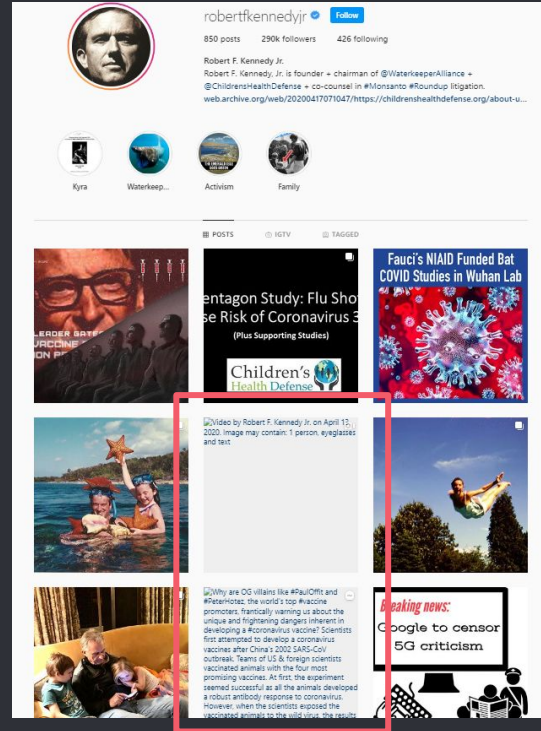
● Quality of the 6 Replayable Mementos

Only 2 Complete, Replayable Mementos

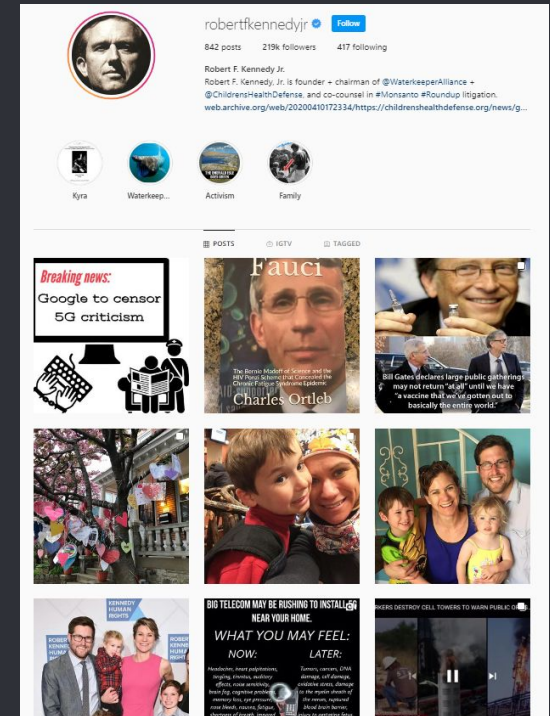
No loaded post images: 3



Partial loaded post images: 1



Complete loaded post images: 2





598

24

6

2



Subtract 574 mementos that redirect to the login page

Subtract 18 revisits

Subtract 4 mementos with missing post images

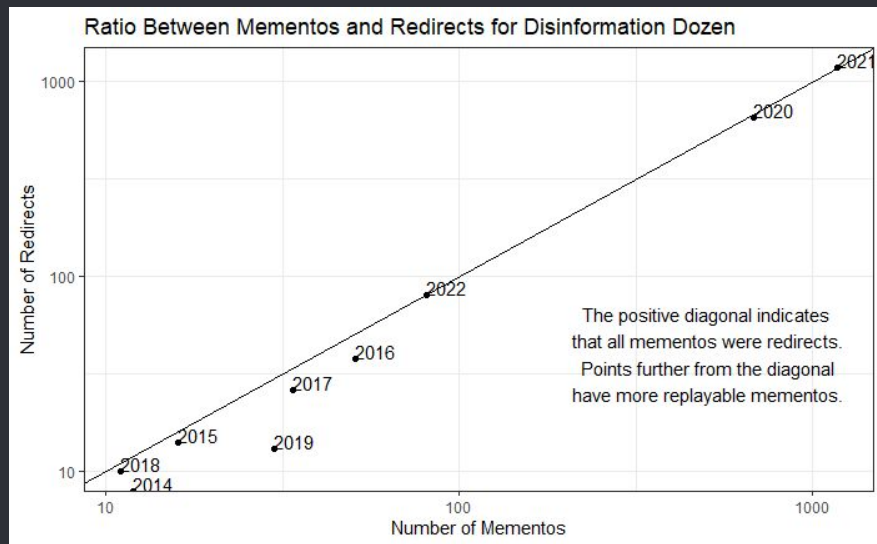
3

How Well is Instagram Archived through the Years?

Trends for the Disinformation Dozen and Health Authorities

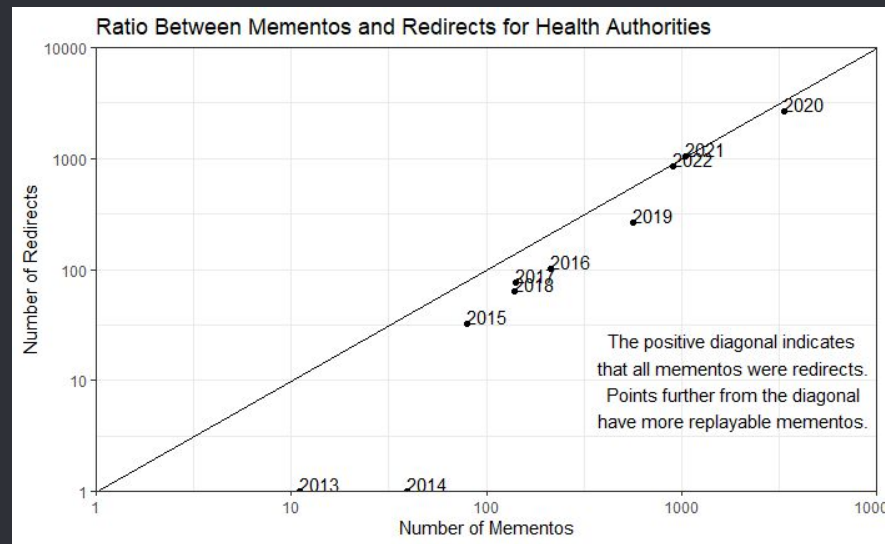
● How Many Mementos Redirected to the Login Page?

Disinformation Dozen



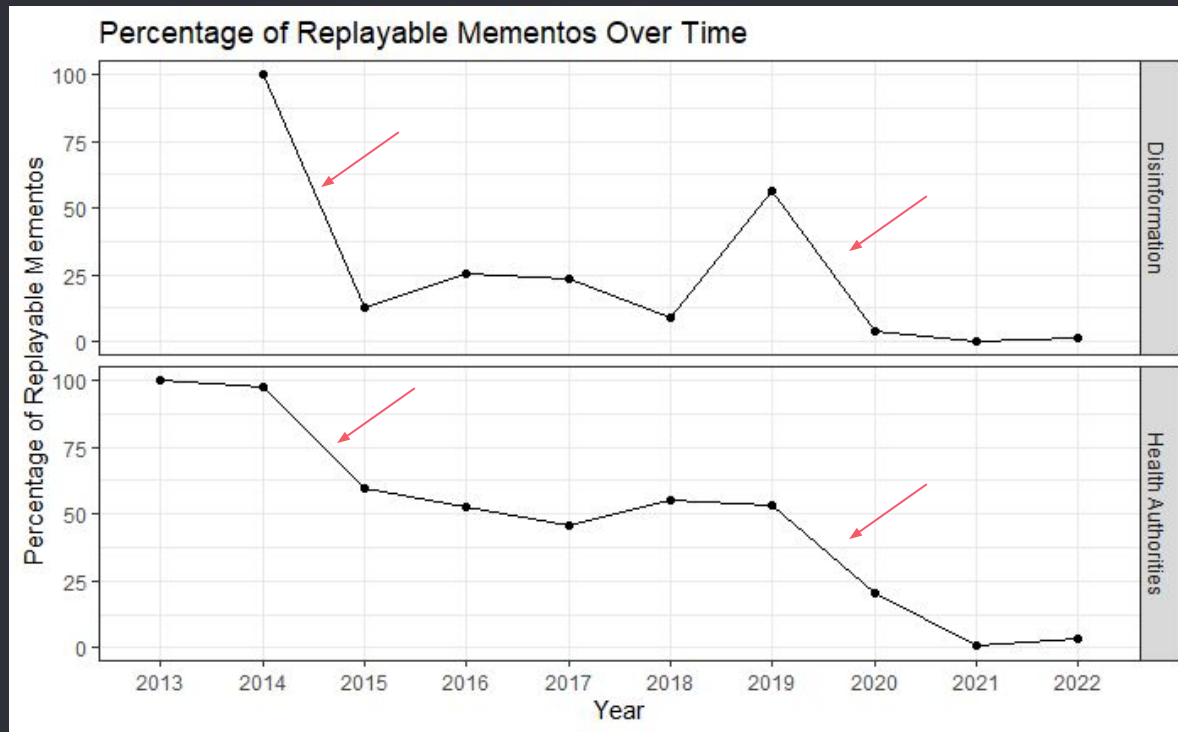
2014 and 2019 as outliers

Health Authorities



Number of replayable mementos decreasing over time

- Decrease in Percentage of Replayable Mementos Over Time



Measuring the Number of Post Images Replayed

The screenshot shows a web browser displaying a social media profile for 'drmercola'. The profile includes a header with a profile picture and a post featuring an image of avocados. Below the header, there are statistics: 137 posts, 2,780 followers, and 0 following. The main content area shows a grid of posts, with the first post also featuring an image of avocados. To the right of the browser window, the Network tab is open, showing a list of resources loaded by the page. The table below represents the data from the Network tab.

Name	Status	Type	Initiator	Size	Time	Waterfall
10691624_802725203084237_5336879...	404	text/html	drmercola	93.4 kB	115 ms	
10727277_753283521405414_1253338...	404	text/html	drmercola	93.4 kB	274 ms	
1391244_1538316683047770_1432973...	404	text/html	drmercola	93.4 kB	491 ms	
10693469_78513884858719_7756324...	404	text/html	drmercola	93.4 kB	156 ms	
10731688_1475024982780340_138657...	404	text/html	drmercola	93.4 kB	184 ms	
10691678_1578526575709532_209144...	404	text/html	drmercola	93.4 kB	204 ms	
925352_708697609225850_159921815...	302	text/plain ...	drmercola	694 B	205 ms	
10683931_784537741593182_1236260...	404	text/html	drmercola	93.4 kB	268 ms	
1724414_1567020900196822_1967016...	404	text/html	drmercola	93.4 kB	234 ms	
10693415_624953277621703_7165317...	404	text/html	drmercola	93.4 kB	254 ms	
1690313_1516191981954817_1021460...	404	text/html	drmercola	93.4 kB	204 ms	
10707128_918152941547159_3650600...	404	text/html	drmercola	93.4 kB	204 ms	
10661266_810267892358442_7187230...	404	text/html	drmercola	93.5 kB	205 ms	
10683888_685341714875976_1081270...	404	text/html	drmercola	93.4 kB	259 ms	

https://web.archive.org/web/20141018180055/http://scontent-b.cdninstagram.com/hphotos-xpa1/t51.2885-15/925352_708697609225850_1599218157_a.jpg 302 ←

https://web.archive.org/web/20141018180055/http://scontent-b.cdninstagram.com/hphotos-xaf1/t51.2885-15/10727277_753283521405414_1253338834_a.jpg 404

https://web.archive.org/web/20141112065738/http://scontent-b.cdninstagram.com/hphotos-xpa1/t51.2885-15/925352_708697609225850_1599218157_a.jpg 200 ←

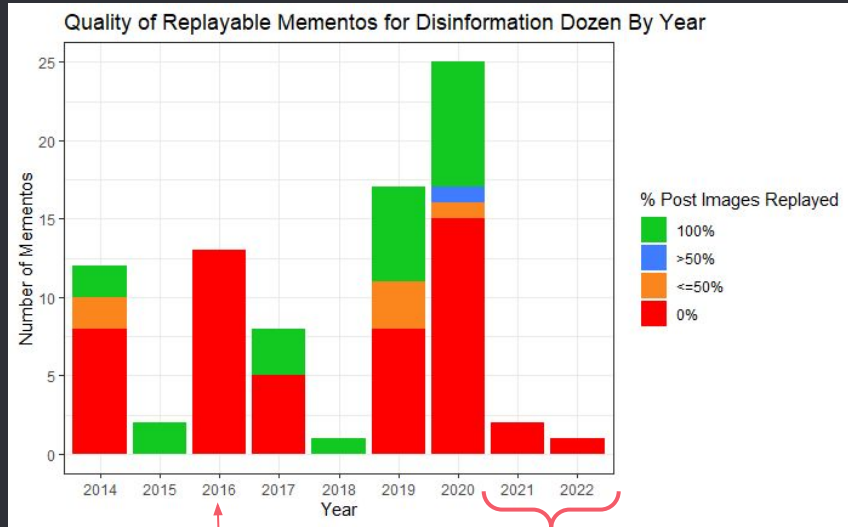
https://web.archive.org/web/20141018180055/http://scontent-a.cdninstagram.com/hphotos-xfa1/t51.2885-15/10691624_802725203084237_533687919_a.jpg 404

{'925352_708697609225850_1599218157-': [302, 200]}, '10727277_753283521405414_1253338834-': [404], '10691624_802725203084237_533687919-': [404],

There is 1 present resource out of 20 total resources for <https://web.archive.org/web/20141018180055/http://www.instagram.com/drmercola>

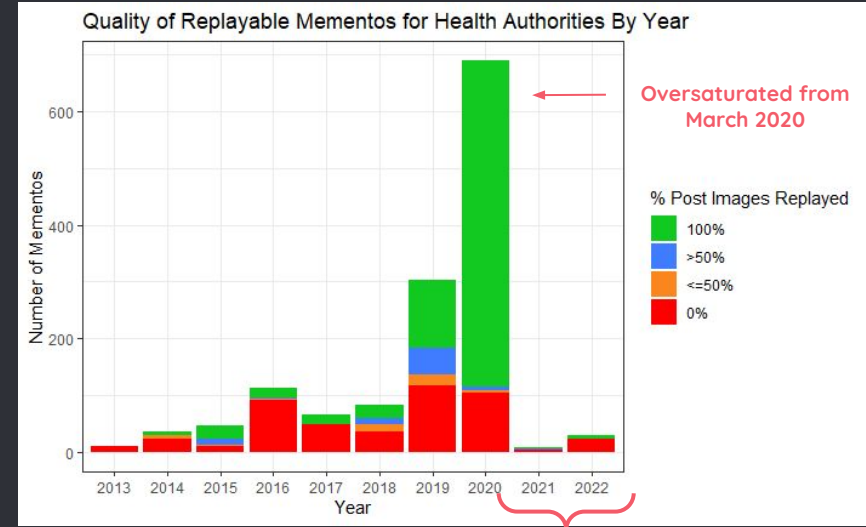
• Many “Replayable” Mementos Are Incomplete

- 64.2% of replayable mementos of Disinformation Dozen Instagram profiles have 0 replayable post images



Particularly low quality

Hardly any replayable mementos



Oversaturated from March 2020

Hardly any replayable mementos

● Future Direction

- Repeat analysis on other groups of disinformation actors

Map tagging networks to identify key players and connections

Study the rise and fall of specific disinformation hashtags over time and their role in spreading malicious content

Look at Instagram UI changes over time to identify when changes occurred and what features make archiving difficult

● Conclusions

- For banned accounts, archives are the only way to study the content that was posted.

Instagram itself is not well-archived, and 96.13% of mementos from Disinformation Dozen accounts redirect to the login page.

The mementos that do not redirect to the login page may not replay every post image. Only 27.16% of replayable mementos for the Disinformation Dozen replay every post image.

Overall, only 1.05% of mementos for the Disinformation Dozen accounts are replayable with complete post images.