

Data Craft: The Manipulation of Social Media Metadata

Author: Amelia Acker

CS 895: Web Archiving Forensics

Presented by Travis Reid

October 29, 2020

Introduction

- Craft designates work that is supplemental, material, and skillful
- Data craft is the type of data craftwork that plays with
 - Platform features
 - Automated operations of platforms
- Media manipulators use
 - Data craft
 - To create disinformation that involves falsified metadata
 - Platform features
 - Spread disinformation
 - To evade moderation efforts that use engagement activities
- Data that involves engagement activities can be read by
 - Machine-learning algorithms
 - Platforms
 - Humans

Manipulation Examples

- Manipulators can generate clicks and fake engagement by
 - Astroturfing
 - Using botnets
- Manipulators can create "click farms" that
 - Promote celebrities
 - Create fake reviews
 - Sell followers and views
 - Promote content
- Political Disinformation

Political Disinformation

- Political disinformation campaigns try to
 - Influence civil discourse
 - Erode democracy by causing mistrust
 - Interfere with elections
 - Hijack news platforms
- These types of campaigns create authentic looking content which makes it hard to detect
- Web archives
 - Can be exploited by manipulators
 - Can be used to trace manipulation campaigns
- When traces are discovered, they disappear from platforms
- There are few collections of disinformation campaigns

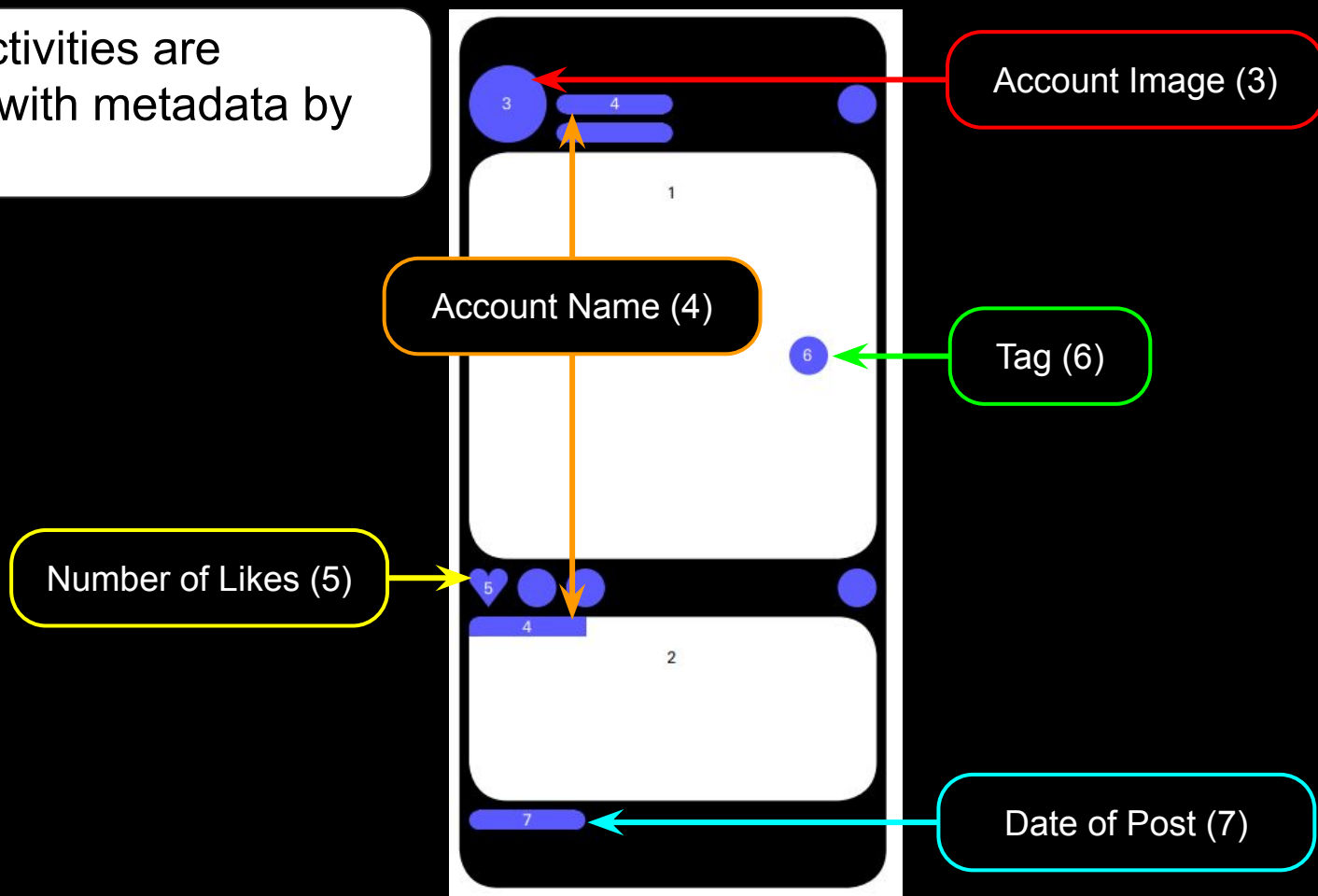
Defining Metadata

- Definition of metadata can change depending on
 - Who is using the data
 - Purposes for using the data
- Metadata is the names that represent aggregated data
 - This definition of metadata
 - Considers the problems with accurately naming data when it is being collected
 - Used in different aggregation contexts
- A difficulty with defining metadata by context is that aggregated data can change over time
- Metadata labels how a digital object is
 - Generated
 - Stored
 - Circulated in networked communication systems
- Difficulties with distinguishing between data and metadata
 - When the owner of data grant new access to it
 - Change the terms of its governance
 - Give it new status or meaning

How Metadata Can Be Used

- Social media metadata can be used as contextual evidence of manipulation
- Metadata can be used as a method to
 - Validate data on social media
 - Help with understanding the craftiness of media manipulators
 - This can help platforms mitigate falsified content
- Metadata with aggregated data can help with:
 - Developing meaning
 - Creating claims
 - Making decisions
 - Creating evidence
- Metadata categories can be used to find, retrieve, or provide new paths to accessing content in platforms
- Other Examples
 - Determining the social media accounts with the most followers
 - Finding the games have been downloaded the most
 - Determining how many times a YouTube video has been watched and shared
 - To target ads on Facebook
 - Restrict options available on a dating platform

All user activities are classified with metadata by platforms



Additional Metadata From API

- location_id
 - Place associated with the post
- profile_views
 - Number of times the account was viewed by other users
- created_time
 - The creation time of the post
- Source
 - The application used to create the post

Platform Activity Signals

- Platform activity signals are social metadata that is about engagement activities
- Social media metadata that are platform activity signals:
 - Username
 - Profile handle
 - Bio field
 - Dates of posted photos
 - Followers and following counts
 - Hearts on posts
- These activities can be read by
 - Machine-learning algorithms
 - Platforms
 - Humans
- Media manipulators use platform features in unexpected ways to avoid moderation efforts that use these metadata categories
- Less skillful manipulators accidentally leave behind digital fingerprints

Digital Fingerprints

- Examples:
 - Accounts with little interaction
 - Inauthentic conversations
 - Review of the account administrators IP addresses from where content was created
 - Locations of most of the account and their followers
 - Type of currency used to buy the ad
 - Paying for ads in rubles (currency used in Russia)
 - Geolocation tags of fake news from a foreign country
- Some fingerprints are only accessible to platform engineers or those with access to the APIs

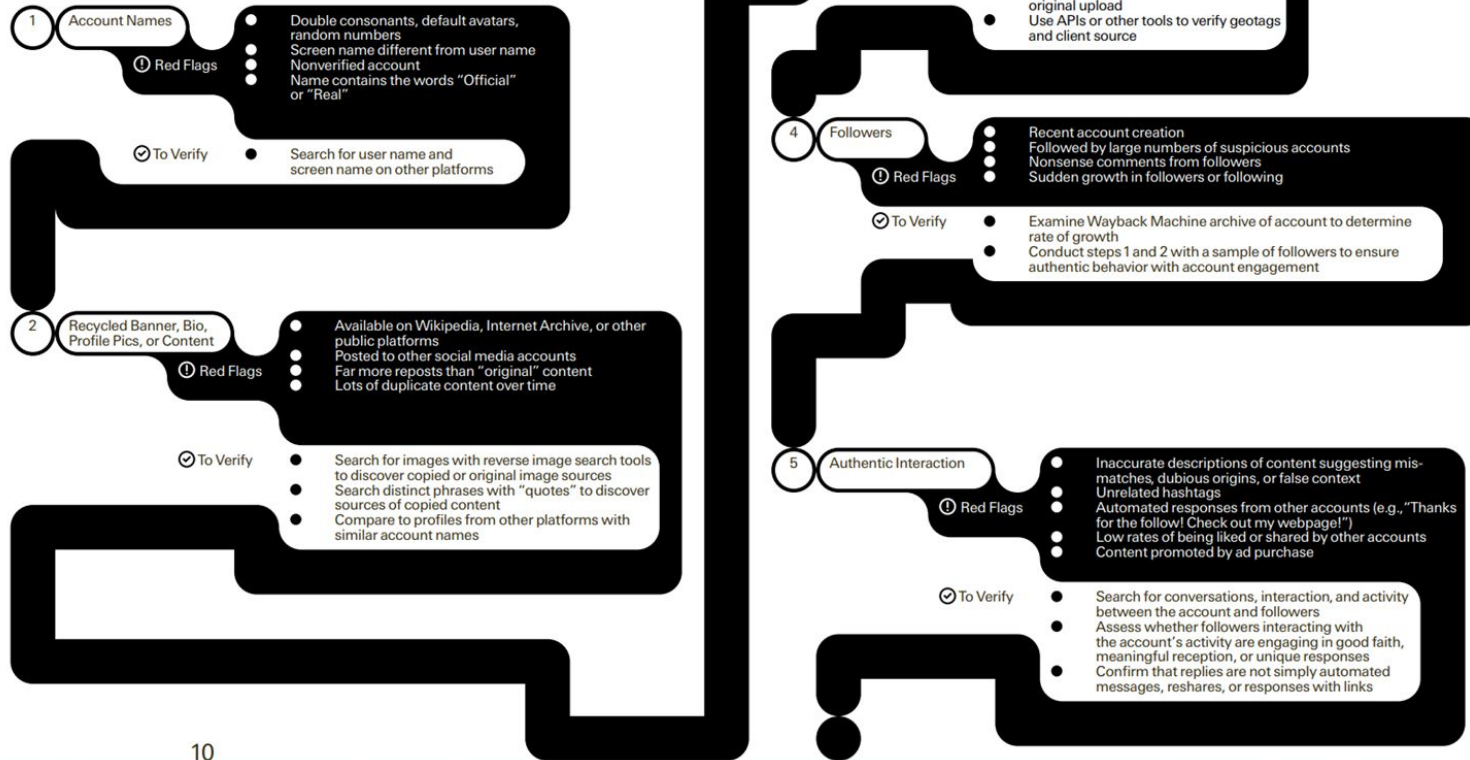
Reading Metadata

- Methods are needed for reading when, where, and how manipulators leverage metadata
- Reading metadata helps with
 - Understanding the craft of data work and the roles of metadata in platforms
 - Identifying vulnerabilities in platforms
- These methods need to account for data craft
 - Craftwork can be read by examining metadata signals generated from adversarial techniques

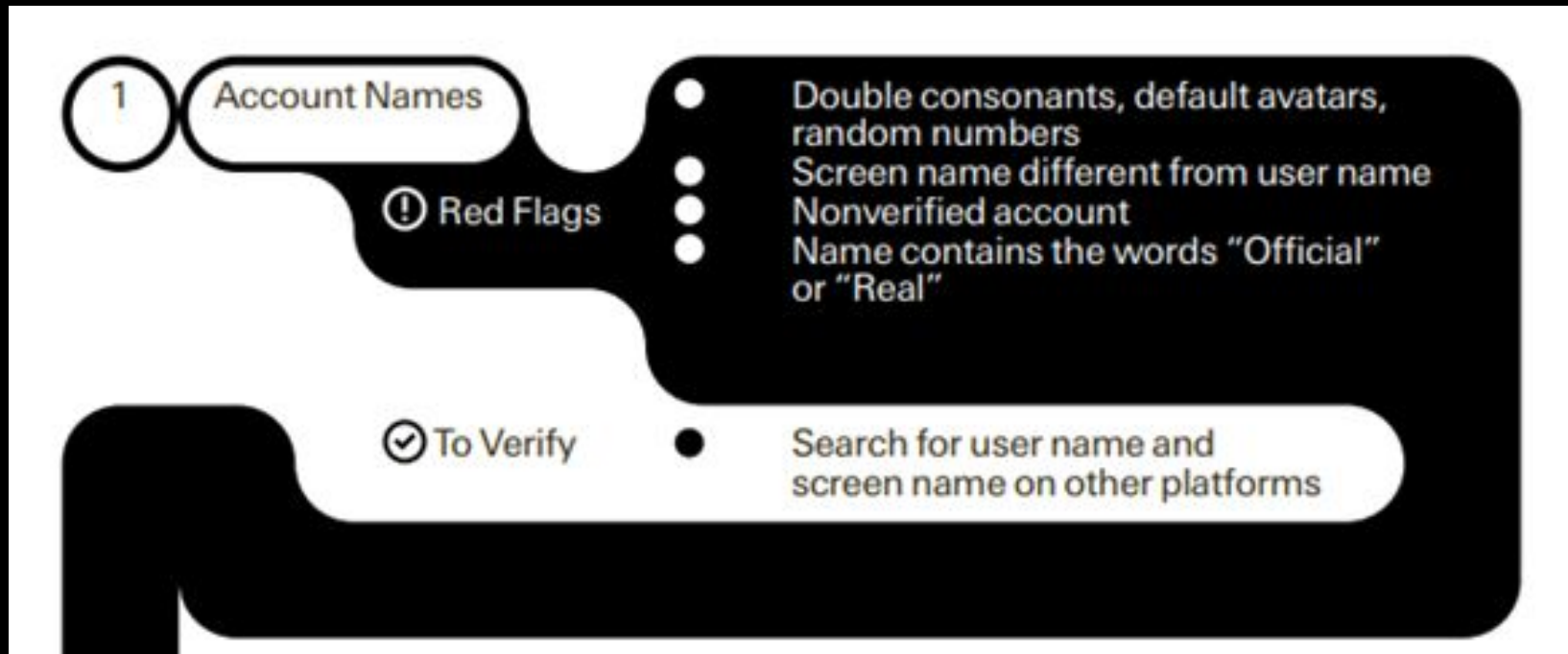
Fig. 2

Reading Metadata

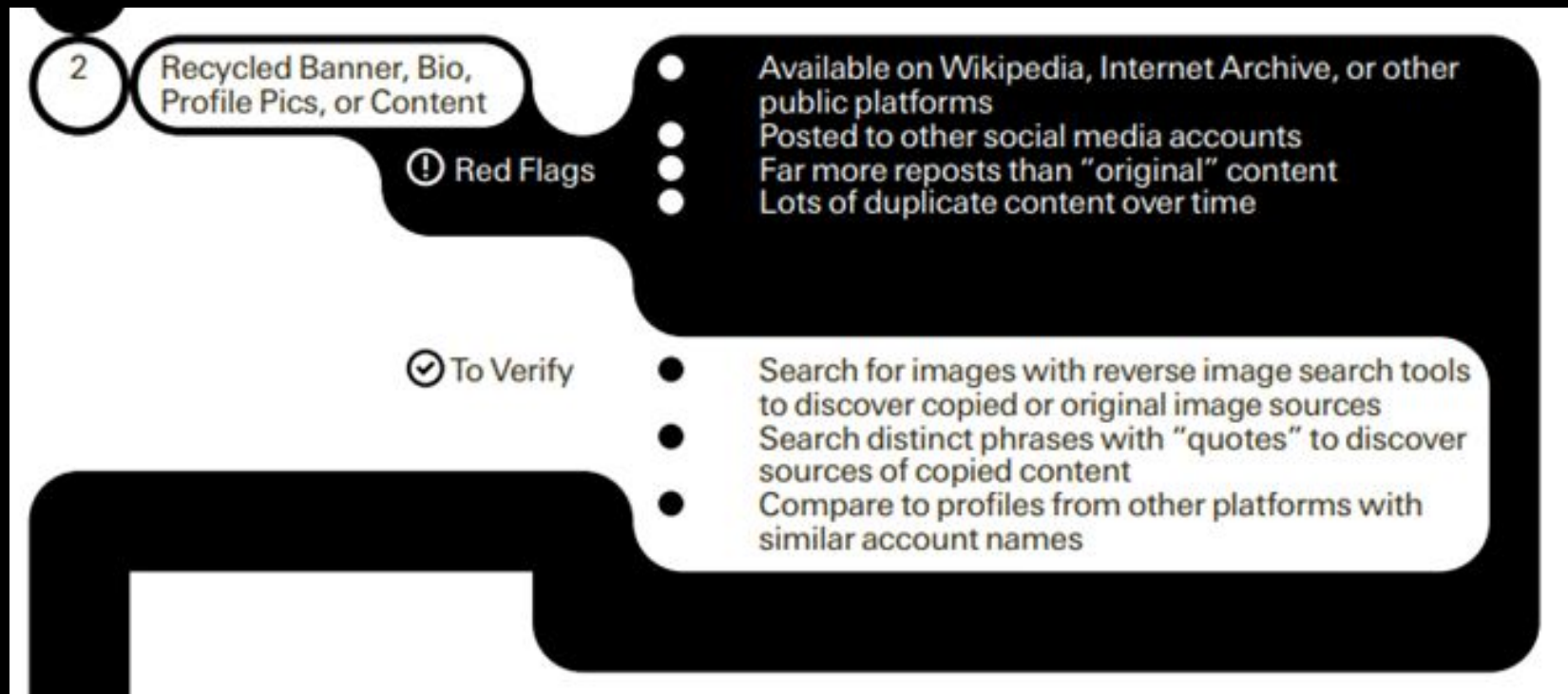
The chart captures a step by step process for reading metadata from social media content. The goal for each step is to evaluate different types of "red flags"—characteristics which can, when taken together, indicate likely manipulation and coordinated inauthentic behavior. None of these red flags can be interpreted as concrete evidence on their own. However, when taken together all of the following metadata categories—including interaction between other accounts—allows readers, researchers, and users to see the traces of manipulative data craft. By examining the interaction between accounts and their followers, steps 4 and 5 allow readers to locate evidence of manipulation and disinformation resulting from coordinated engagement strategies that generate inauthentic behavior.



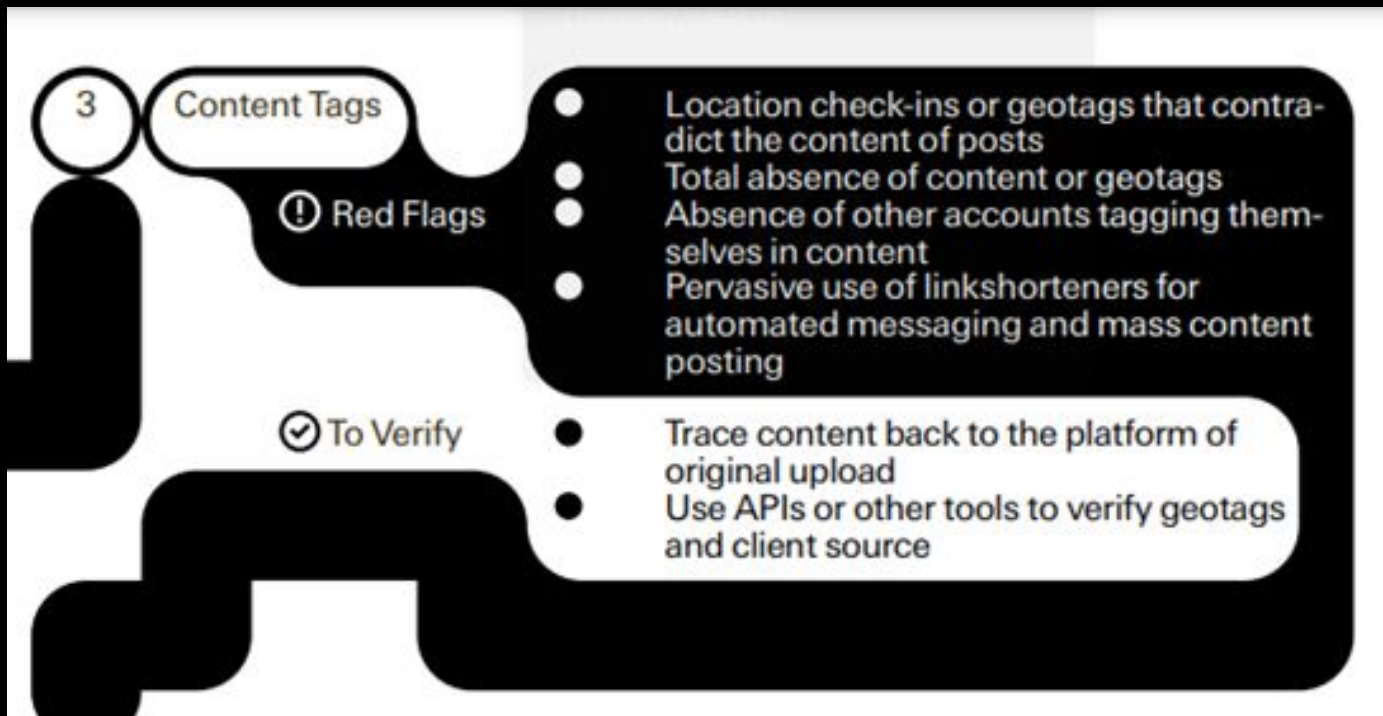
Check for errors in account name and verify the username across different platforms



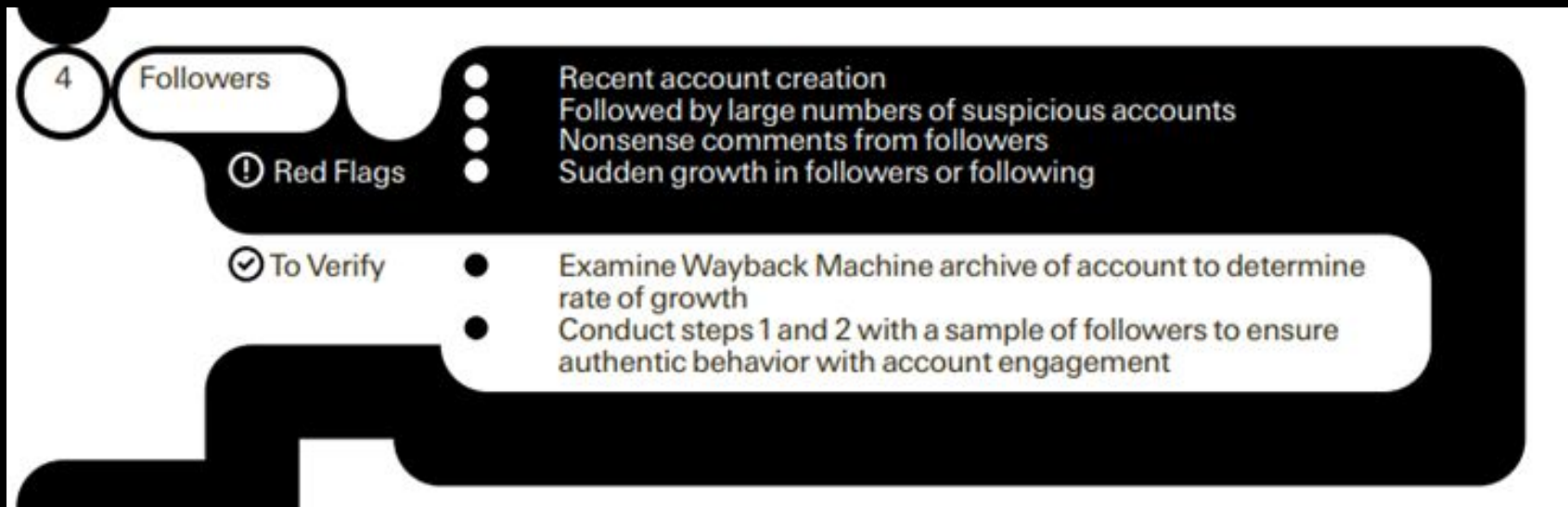
Check to see how often the account creates original content



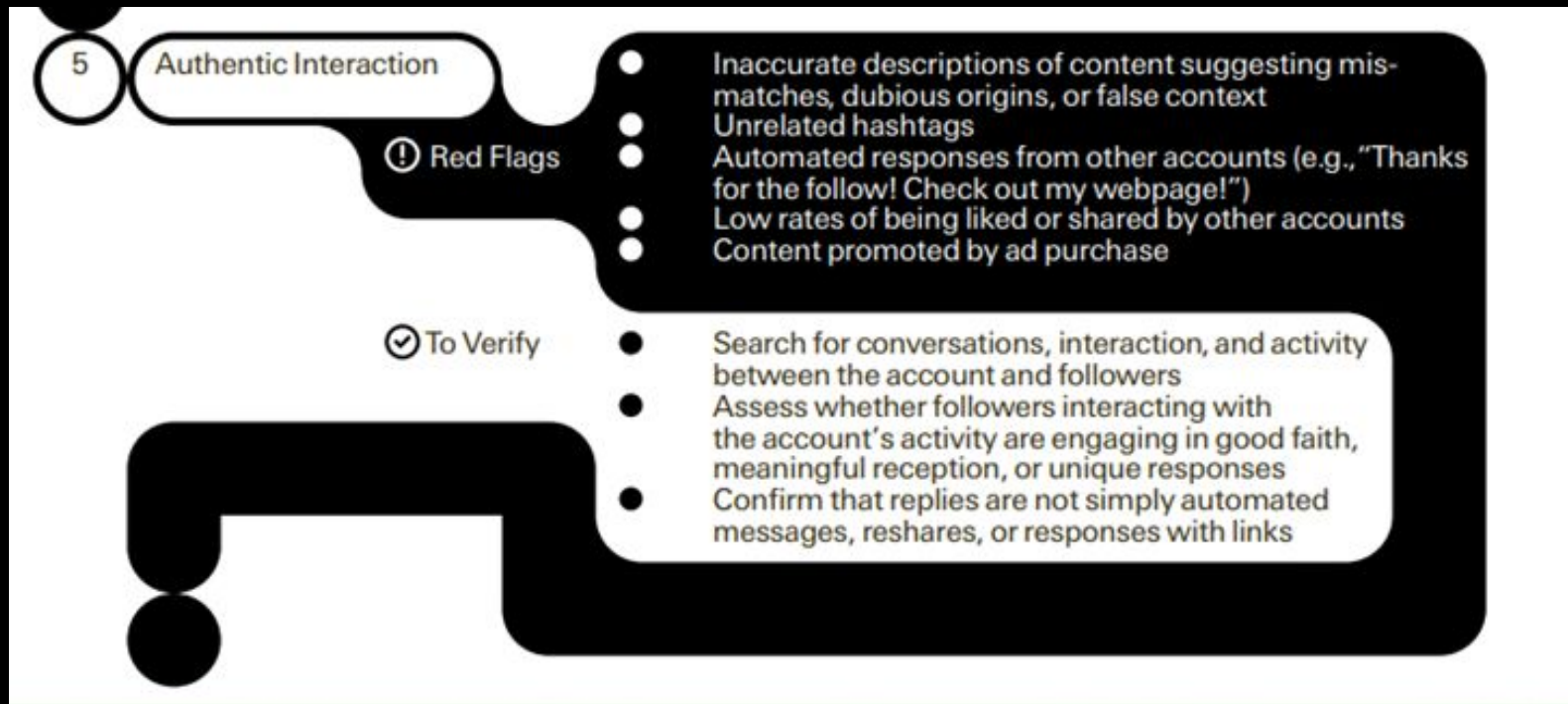
Look for the use of link shorteners and the absence or contradiction of content and geotags



Determine if the followers of the account are authentic by checking the comments and seeing if there is a sudden growth of followers do not engage with the account



Confirm the authentic interaction between the account and followers by checking if the followers' reply is related to the content, does not seem like an automated response, and the content is not a promoted ad



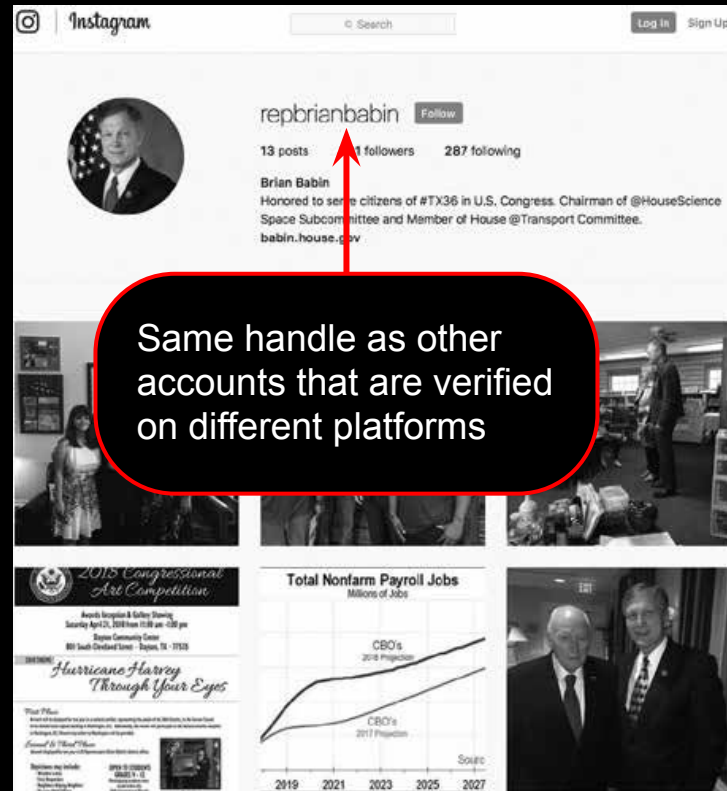
Case Studies

1. Babin on Instagram: Mimicking Legitimacy
2. Creating Imposter Accounts: Claiming Deleted Screen Names
3. Facebook Internet Research Agency (IRA) Ads

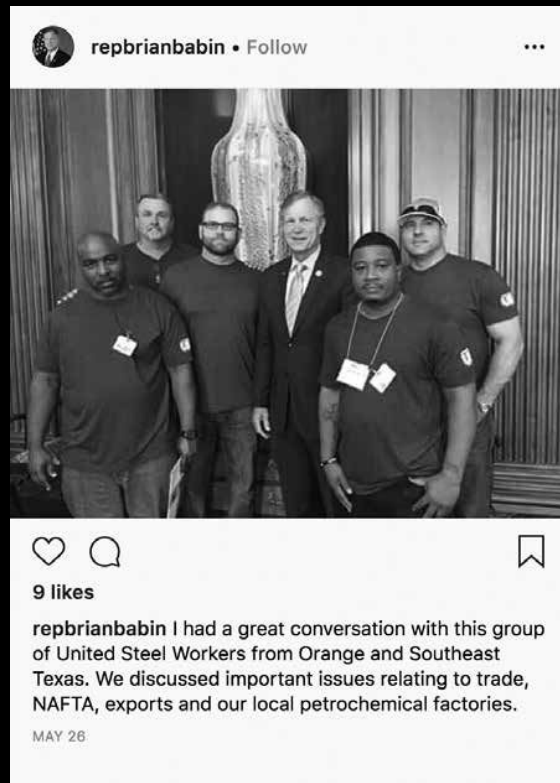
Babin on Instagram: Mimicking Legitimacy



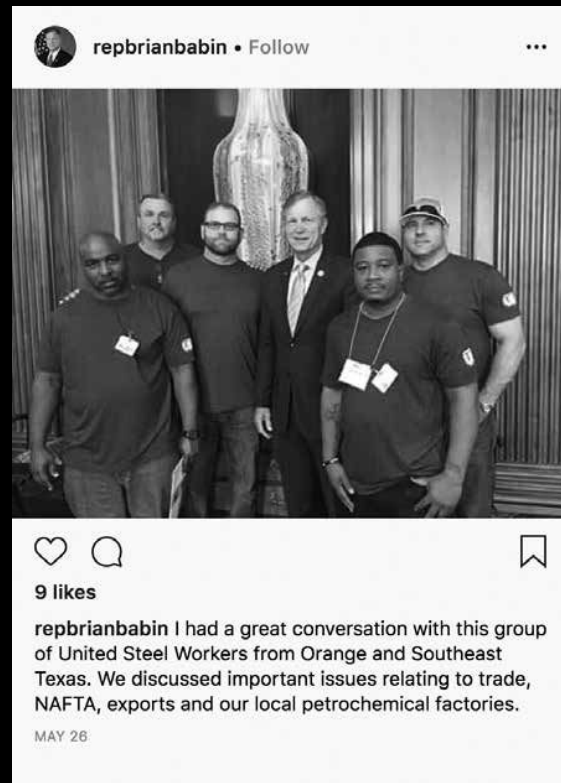
Babin on Instagram: Mimicking Legitimacy



Post Comparison



Post Comparison



Which Account Is Legitimate?

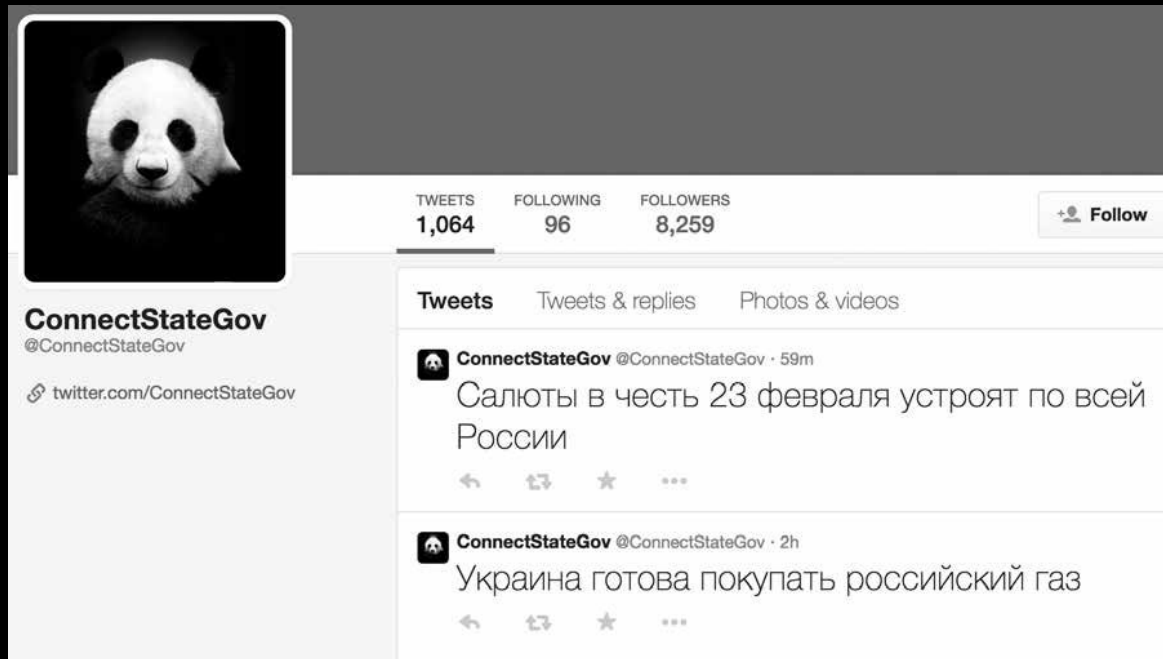
- @babin.official account is more suspicious
 - Few posts have comments and likes
 - Duplicate posts have newer time stamps than @repbrianbabin's posts
 - These activity signals are indicators of how context can be mimicked, gamed, or falsified
 - @repbrianbabin account has been tagged by legitimate users
- Automated moderation would be difficult for accounts like @babin.official
 - Small number of followers and followed accounts
 - Not a lot of engagement or activity

Tips For Determining The Legitimate Account

- Across different platforms
 - Search for accounts with a similar
 - Real name
 - Username
 - Account handle
 - Compare
 - Profile pics and Account Banners
 - Perform a reverse Google image for these images
 - Account descriptions
 - Identify copies of posts and the timestamps
 - Find absences that have not been copied or backfilled
 - Lack of comments
 - Post description
- Check if the account has been tagged by verifiable platform users
- Consider parody or organizational change before assuming that something is malicious manipulation

Creating Imposter Accounts: Claiming Deleted Screen Names

- Manipulators can accomplish a full account takeover for a legitimate account
- Account metadata may look authentic because of previous activity, but activity signals may be different
- Littman noticed that several US government accounts were tweeting in Russian
- This is a vulnerability with Twitter screen names which allows imposter accounts to claim the screen name of a deleted account
- Accounts are removed after no activity for 6 months



U.S. Digital Registry

- Digital.gov has a reference database that can be used to help to confirm the official status of social media accounts
 - U.S. Digital Registry: <https://digital.gov/services/u-s-digital-registry/>
- Littman found 100 deleted accounts and 29 suspended accounts that were listed as active official government accounts by the U.S. Digital Registry

Littman Performed A Full Account Takeover

- Littman performed an account takeover for the official @USEmbassyRiyadh account
 - He waited for the account to be deleted before claiming the screen name
 - Copied the previous account metadata from Wayback Machine web archive
 - Profile image and banner image
 - Name
 - Location
 - Other metadata
 - Tweeted a Wilford Brimley quote
 - Archived the page in the Wayback Machine's web archive
- This example shows how web archives can be exploited by manipulators



Tips For Detecting Imposter Accounts

- Determine when the account was created
- Look at the number of tweets/posts that have been created since the account start date
- Look for duplicate attached media (pictures, videos, links)
- Check if the account has been dormant by viewing the date of the last post or activity
- For accounts listed as an official account, confirm the account by visiting the personality/institutional home page
- Search web archives like Internet Archive's Wayback Machine for crawls of the account
 - View multiple crawls to see if the account was dormant or has ever been deleted or suspended
- Perform a reverse Google image search for the profile pics and banners to see if there are copies
- Determine the authenticity of the followers and commenters
- Check to see if the comments are substantive and engaging with the content

Facebook Internet Research Agency (IRA) Ads

During 2018, 3,517 Facebook and Instagram ads were published by the Internet Research Agency which is a Russian propaganda firm

- Ads transmitted in zipped PDF files
 - This is a version of data craft that is used to prevent others from analyzing trends
 - PDF format is difficult to extract structured data from
- This ad appeared across YouTube, Facebook, and Twitter
- William&Kalvin account
 - Bought promotional ads
 - Had little interaction with other users
 - Pushed frequent video content
 - Initial content on police brutality and racism
 - During 2016 posted anti-Clinton content with conspiracy theories



Williams&Kalvin Sponsored · 

 Like Page

Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and our judicial system is absolutely blind. We are all Americans, but why does our corrupt Government differ black and white people? We want the same attitude! I don't want to be scared of living in my country! They will never shut me up!

<https://www.youtube.com/watch?v=GjSy0RgShOI>

 **Police is not above the law!**

Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and our judicial system is absolutely blind. We are...

YOUTUBE.COM

58 Likes · 3 Comments · 21 Shares

 Like  Comment  Share

Extracting Metadata From The PDFs

- Ed Summers created software to extract the images and metadata from the PDFs
 - Simon Willison wrote software to convert the output to a searchable database (<https://russian-ira-facebook-ads.datasettes.com/>)

Database Entry ID And Image

id
img

789

 Williams&Kalvin
Sponsored · 

 Like Page

Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and our judicial system is absolutely blind. We are all Americans, but why does our corrupt Government differ black and white people? We want the same attitude! I don't want to be scared of living in my country! They will never shut me up!

<https://www.youtube.com/watch?v=GjSy0RgShOI>



Police is not above the law!

Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and our judicial system is absolutely blind. We are...

YOUTUBE.COM

58 Likes · 3 Comments · 21 Shares

 Like  Comment  Share

IRA Ad Targeting Metadata

targeting location:United States, age:18–54, language:English (UK), language:English (US), placements:News Feed on desktop computers, placements:News Feed on mobile devices, accessing_facebook_on:Wi-Fi, people_who_match:interests:BlackNews.com, people_who_match:interests:HuffPost Politics, people_who_match:interests:HuffPost Black Voices, and must also match:behaviors:African American (US)


- Proxies for specific targeting
 - Associated interests
 - Partisan media
 - Device choices
- Facebook moderation previously relied on users to flag content
- The review process is more strict now to prevent violations of Terms Of Service

Ethnic affinity categories are gathered from outside of Facebook platform

- Ethnic affinity categories were being used to target ads to or away from users
 - Prevent housing ads from being seen by African Americans or Asian Americans
 - Promote content to users with anti-Semitic interests
- Facebook disabled the use of ethnic affinity marketing for certain ads

Other Metadata For The IRA Ad

impressions	3067
clicks	172
url	https://www.youtube.com/watch?v=Gij0RgShO
text	Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and ourjudicial system is absolutely blind. We are all Americans, but why does our corrupt Government differ black and white people? We want the same attitude! I don't want to be scared of living in my country! They will never shut me up! https://www.youtube.com/watch?v=Gij0RgShO Where is the justice? Our brothers and sisters are being cruelly killed by the so-called police every day and ourjudicial system is absolutely blind. We are... Police is not above the law!
spend_usd	16.0
spend_amount	1000.00
spend_currency	RUB
created	2016-01-05T02:04:48-08:00
ended	2016-01-07T02:03:00-08:00

 Russian Currency (Ruble)

Remaining Sections Of Database Entry

Advanced export

JSON shape:

☐ Default

☐ Array

CSV Options:

☐ Download File

☐ Export CSV

```
CREATE VIEW display_ads AS
select ads.id,
       case when image is not null then
         json_object("img_src", "https://raw.githubusercontent.com/edsu/irads/03fb4b/site/" || image, "width", 200)
       else
         "no image"
       end as img,
       json_group_array(
         json_object(
           "label", targets.name,
           "href", "/russian-ads/display_ads?target="
             || urllib_quote_plus(targets.id)
         )
       ) as targeting,
       ads.impressions, ads.clicks, ads.url, ads.text,
       cast(case
         when ads.spend_currency == "RUB" then ads.spend_amount * 0.016
         else ads.spend_amount
       end as float) as spend_usd,
       ads.spend_amount, ads.spend_currency,
       ads.created, ads.ended
from ads
join ad_targets on ads.id = ad_targets.ad_id
join targets on ad_targets.target_id = targets.id
group by ads.id
order by ads.id
```

Dataset Entry:

https://russian-ira-facebook-ads.datasettes.com/russian-ads-919cbfd/display_ads?_search=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DGij0RgShO%27C&id_exact=789

Tips For Reading Metadata

- Use social dashboards to see
 - Account creation time and date
 - Average daily active posts
- Examine
 - Promoted posts
 - Ads policies of platforms
- Consult the page administrator's
 - User account and page
 - Compare the rate of posting promoted content to free content
- Examine how often content is shared
 - Does the post include the same memes or videos but with different captions?

Conclusions

- Data craft is the type of data craftwork that plays with platform features and automated operations of platforms
- We can identify and read user generated contextual metadata and additional metadata fields
- Reading metadata serves as a method for identifying disinformation
- Media manipulators understand the problems with using automated moderation techniques that relies on platform activity signals
 - Their craftiness can create illegitimate data that is considered authentic data
 - They also know how to target specific users with ad technology
- Web archiving can be used to trace manipulation campaigns which is useful, because disinformation may be deleted after it is identified
- The case studies this report has shown how reading metadata can help with understanding the craft of data work and the roles of metadata in platforms