

## A Security Analysis of Facebook's Political Ad Library

Laura Edelson 11/13/2019

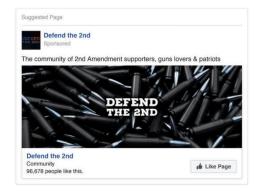


#### Background

### Russian Disinformation on Facebook during 2016 election











### Russian Disinformation on Facebook during 2016 election







93 Reactions 6 Comments 20 Shares



#### Restrictions on Political Advertisers

- Facebook, Google, and Twitter begin some form of disclosure of payers for political advertising
- Major platforms began some form of verification of identity of political advertisers
- Major platforms place restrictions on who can pay for political advertising, usually geographic

#### Transparency Archives for Political Content

- Facebook Ad Archive launched May 7<sup>th</sup>, 2018
- Google Political Ad Transparency dataset launched May 31st, 2018
- Twitter Ad Transparency Center launched June 28th, 2018



#### **Facebook**

- Platforms: Web Portal and API
- Accessible to any ID verified user who agrees to Terms of Service
- Data available:
  - Ad Creative Text/Image/Video
  - Impression & Spend (Bucketed)
  - Demographic & Geographic distribution of impressions
  - Advertiser-specified start date, end date, disclosure string
  - Page that the ad ran against



#### Google

- Platforms: Web Portal and Big Query dataset
- Accessible to anyone, no login required
- Data available:
  - Link to composed Ad for non-Third Party ads
  - Impression & Spend (Bucketed)
  - Demographic & Geographic targeted groups
  - Start & End date
  - Vetted Advertiser identifying information (FEC id, EIN)



#### **Twitter**

- Platforms: Web Portal only
- Accessible to anyone, no login required
- Data available:
  - Ad Creative Text/Image/Video
  - Exact Impression & Spend
  - Demographic & Geographic distribution of impressions
  - · Start & End date
  - · Billing information for Ad payer
- No longer accepting political advertising as of Nov. 22<sup>nd</sup>, 2019



#### Facebook:

\$888M since May, 2018 in the United States

#### Google:

\$126M since May, 2018 in the United States

#### **Twitter:**

<\$5M since June, 2018 in the United States</li>



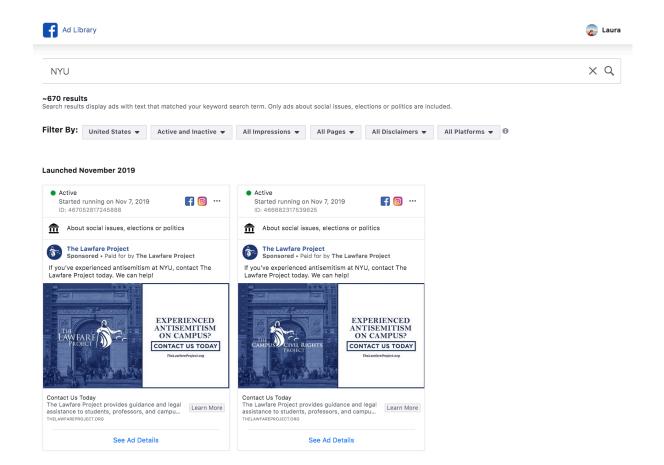
#### Facebook Ad Archive Overview



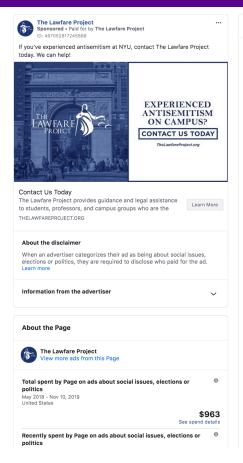
#### In the United States, advertisers are required to

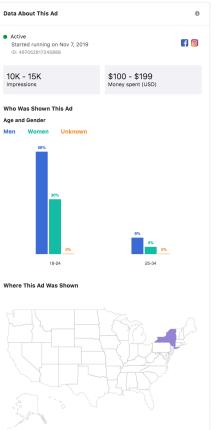
- Disclose ads about "Social Issues, Elections, or Politics"
- Provide an accurate disclosure string naming the entity that paid for the ad, conforming to a standard format



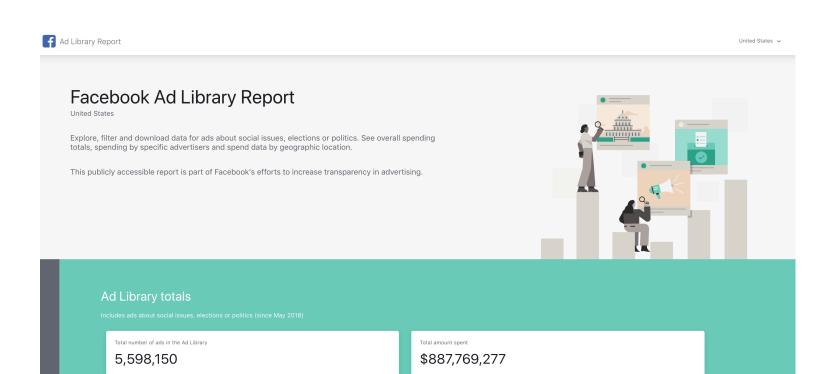














#### Facebook Ad Library API gives access to the same dataset as the Web Portal

- Searchable by keyword string or page id
- No direct access to images or videos, but these can be collected from a link provided by the API
- Most data provided through the Web Portal is available through the API, with the exception of Active status and whether the ad has been removed for a policy violation



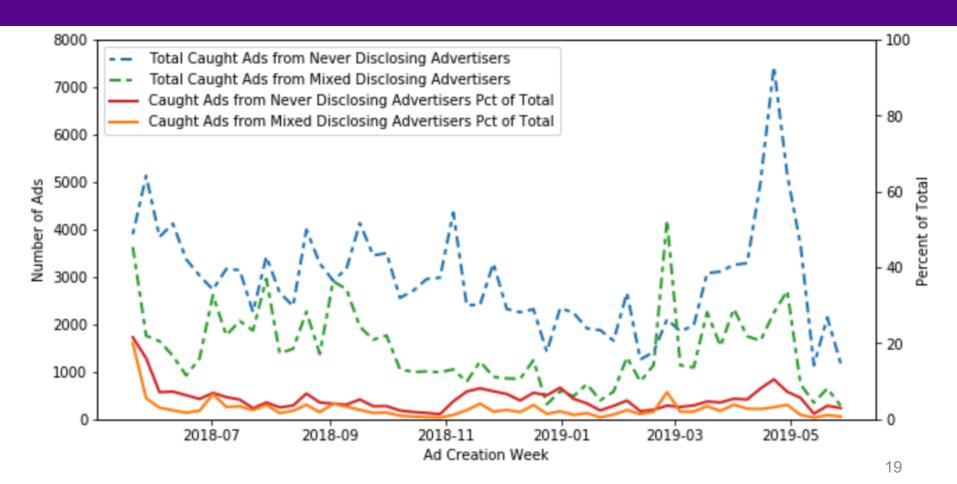
## Security of Facebook's Ad Library



Question 1)
How good is Facebook at enforcing their disclosure policy?



# Question 1) How good are advertisers at voluntarily disclosing ads that meet the criteria for inclusion in the Ad Library?





#### **Question 1)**

How good are advertisers at voluntarily disclosing ads that meet the criteria for inclusion in the Ad Library?

#### **Answer:**

They've gotten a bit better, but there is a problem with persistent non-disclosure



# Question 2) How useful are disclosure strings for understanding which entities are spending money on political advertising?



	Pages (Pct)	Ads (Pct)	Spend (Pct)
Never attributed	69,072 (54.6%)	200,751 (3.2%)	\$15,266,952 (2.4%)
Misattributed Undisclosed	17,297 (13.7%)	156,348 (4.2%)	\$21,996,150 (3.5%)
Misattributed Typo'd	1,776 (1.4%)	299,807 (8.1%)	\$61,019,010 (9.8%)
Total Misattributed	87,926 (69.6%)	655,619 (17.7%)	\$98,153,857 (15.8%)

TABLE VI: Corrected Aggregated Advertiser statistics



# Question 2) How useful are disclosure strings for understanding which entities are spending money on political advertising?

**Answer:** 

They... aren't.



Question 3)
Can we use the information that has been made transparent to find violations of Facebook's policies?



### **Undisclosed Coordinated Activity**



#### What is Undisclosed Coordinated Activity?

Advertisers are required to truthfully disclose who pays for ads that meet the criteria for inclusion in the Ad Library. Facebook has periodically banned advertisers who engage in "Coordinated Inauthentic Behavior"



#### **Example of disclosed coordinated activity:**











**Target:** 

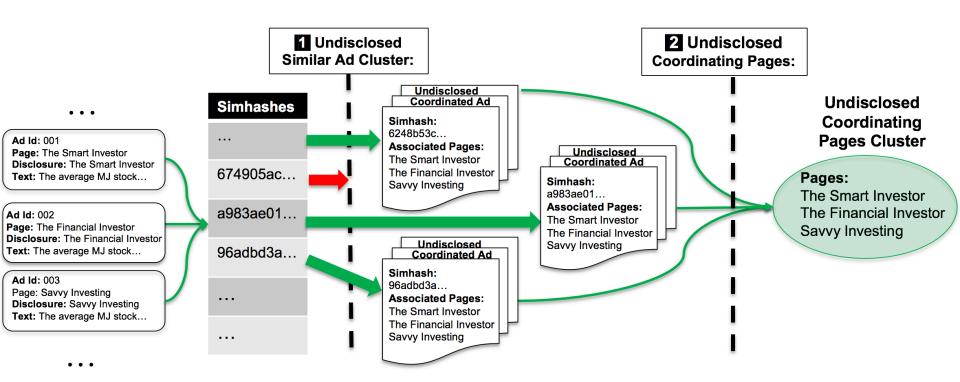
**Undisclosed Coordinated Activity** 

Tool:

Simhashing: A locality sensitive hashing technique that creates hash values such that the difference between two hashes is equivalent to the Hamming distance of the two texts the hashes were created from



#### **Target: Undisclosed Coordinated Behavior**





#### Step 1:

Cluster near-identical ads using Simhashing. Determine mean number of disclosure strings per ad cluster. Establish cutoff for outlier clusters of z-score=2 or more.

#### Step 2:

For ad clusters associated with more than one page, what is mean number of times 2 pages are associated with each other? Establish cutoff for page clusters of z-score=2 or more

Step 3: Identify pages that ran ads in outlier ad clusters. Group those pages together if they also meet the cutoff for page clusters



#### **Corporate Astroturfing**









#### **Dubious Commercial Activity**









#### **Clickbait**

Inactive Started running on May 22, 2018

This ad ran without a disclaimer. (1)



Housekeeping 101 Sponsored

Internet destroys white woman who called the cops on black folks having a BBQ with 10+ brutal memes:



Internet Destroys White Woman Who Called The Cops On Black Folks Having A BBQ With 10+ Brutal Memes

Merritt Lake in Oakland, California, has seen some crazy stuff. According to locals, creepy men are jogging there in Speedos... BOREDPANDA.COM

See Ad Details

Inactive

Started running on May 22, 2018



This ad ran without a disclaimer.



Rick Lax Sponsored

Internet destroys white woman who called the cops on black folks having a BBQ with 10+ brutal memes:



Internet Destroys White Woman Who Called The Cops On Black Folks Having A BBQ With 10+ Brutal Memes

Merritt Lake in Oakland, California, has seen some crazy stuff. According to locals, creepy men are jogging there in Speedos... BOREDPANDA COM

See Ad Details

Inactive

Started running on May 22, 2018



This ad ran without a disclaimer. (1)



**Alonzo Lerone** Sponsored

Internet destroys white woman who called the cops on black folks having a BBQ with 10+ brutal memes:



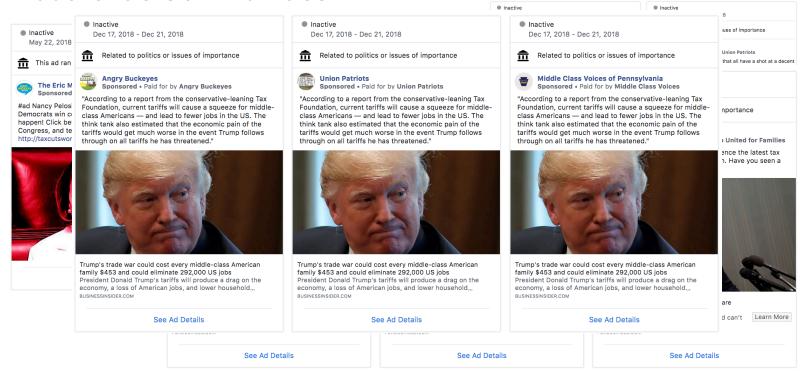
Internet Destroys White Woman Who Called The Cops On Black Folks Having A BBQ With 10+ Brutal Memes

Merritt Lake in Oakland, California, has seen some crazy stuff. According to locals, creepy men are jogging there in Speedos... BOREDPANDA COM

See Ad Details



#### **Inauthentic Communities**





#### **Future Work**



- Cluster Ads by campaign/central idea
- Extract more metadata
  - Named Entity Extraction
  - Sentiment scoring
  - Topic Modeling
- Use clustering and feature extraction to build a screening system for disinformative content



#### **Conclusions**



- Transparency can work!
- Tactics to spread disinformation are still evolving and we are very early in the lifecycle of this behavior
- All parties need to invest more time and money in solving this problem: Uses, Platforms, and Researchers