You Won't Believe It:
Exploring the Advertising Ecosystem of Fake News Websites

Catherine Han
Allen Tong
CS 261N

1

Misinformation Flavors: **Deception**

REUTERS

Russia deploying coronavirus disinformation to sow panic in West, EU document says

Robin Emmott

5 MIN READ

Misinformation Flavors: **Junk Science**

NATURAL NEWS
Defending Health, Life and Liberty 🇺🇸

April 15, 2020 - Ethan Huff

In 1955, a "life-saving" polio vaccine sickened and fatally paralyzed countless children after being declared "safe and effective"

(Natural News) Unless you're of the appropriate age, you probably won't recall firsthand the tainted polio vaccines from back in the 1950s that injured and killed countless children, all in the name of "saving lives." But you should brush up on that history regardless, because it's happening all over again with the Wuhan coronavirus … [Read More...]

How 5G alters blood cell permeability, amplifying coronavirus fatalities

Misinformation Flavors: **Unreliable Clickbait**

BREITBART

FAUCI OKAY WITH TINDR HOOKUPS

Dr. Anthony Fauci told listeners on Snapchat's *Good Luck America* on Tuesday that meeting up with someone after connecting on a dating app is a matter of deciding how much risk one feels comfortable taking.

Misinformation Related Work

- Detection

- Fact-checking

- Prevention

- Propagation modeling

Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations

KATE STARBIRD, Human Centered Design & Engineering, University of Washington
AHMER ARIF, Human Centered Design & Engineering, University of Washington
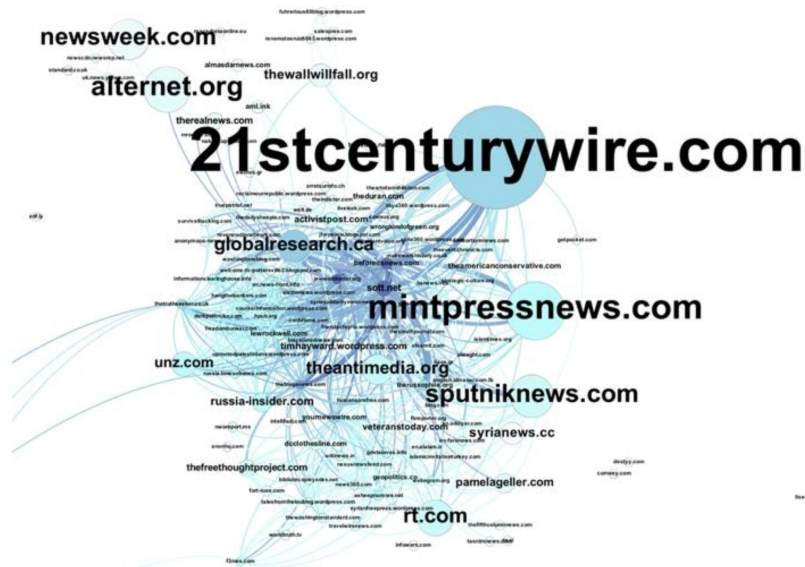TOM WILSON, Human Centered Design & Engineering, University of Washington

Fig. 4. Domain Network Graph: Content Sharing across the "Alternative" Media Echo-System. Each node is a website that hosted an article cited in the White Helmets tweets. View is focused on domains that shared anti-White Helmets content. Nodes are connected when the same article appeared on both domains. Nodes are sized by the number of tweets citing that domain.

5

Online Advertisement Problem Space

- Big revenue stream for "free" services
  - For news publishers too
- Impact on user experience and privacy
  - Annoyance
  - Performance
  - Trackers
  - Fear

Your Ad

CNN

Uh oh. Looks like you're using an ad blocker.

We charge advertisers instead of our audience. Please whitelist our site to show your support for CNN.com
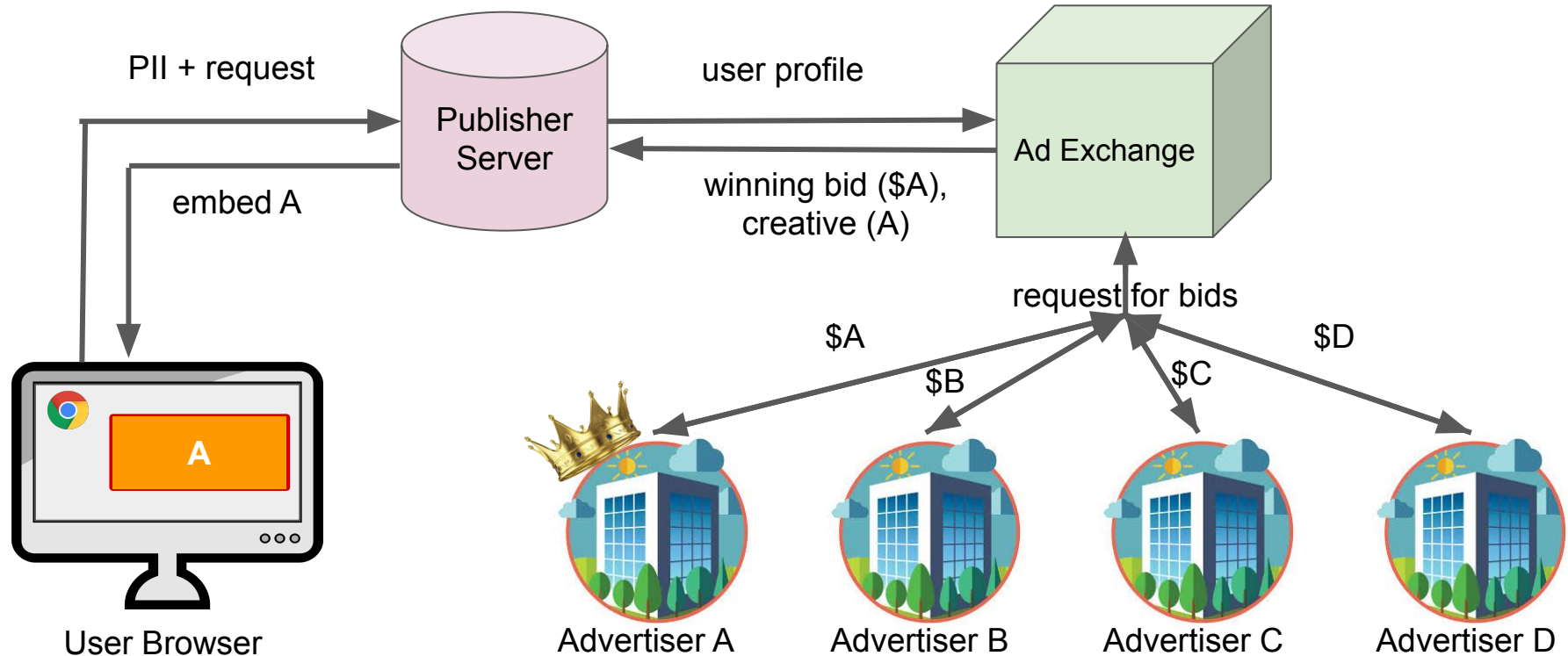
whitelist us

Online Advertisement Problem Space, con't

- Targeting algorithms
  - Algorithmic bias

- RTB ecosystem

Amit Datta*, Michael Carl Tschantz, and Anupam Datta

Automated Experiments on Ad Privacy Settings

seemingly discriminatory ads. In particular, we found that visiting webpages associated with substance abuse changed the ads shown but not the settings page. We also found that setting the gender to female resulted in getting fewer instances of an ad related to high paying jobs than setting it to male. We cannot determine who

Real-Time Bidding (RTB) Simplified

PII + request

Publisher Server

user profile

Ad Exchange

embed A

winning bid ($A), creative (A)

request for bids

$A

$B

$C

$D

User Browser

A

Advertiser A

Advertiser B

Advertiser C

Advertiser D

Misinformation and Ad Revenue

INSIDE THE MACEDONIAN FAKE-NEWS COMPLEX

WIRED

Between August and November, Boris earned nearly $16,000 off his two pro-Trump websites. The average monthly salary in Macedonia is $371.

Misinformation and Ad Revenue, con't

Ceselkoski built seven or eight websites—all oriented toward the American reader. This made sense. In web-advertising terms, an American click is roughly three times more valuable than a non-American click.

Misinformation and Ad Revenue, con't

websites. Facebook and Google have unveiled new systems for screening out misinformation, but they're not built for catching every low-level fib circulating around the internet. Boris won't focus on political fake news, in all probability—but there are plenty of other topics of interest, plenty of websites from which to swipe content, and plenty of potential readers around the world who may click in sufficient numbers to finally buy him his BMW. 🎟
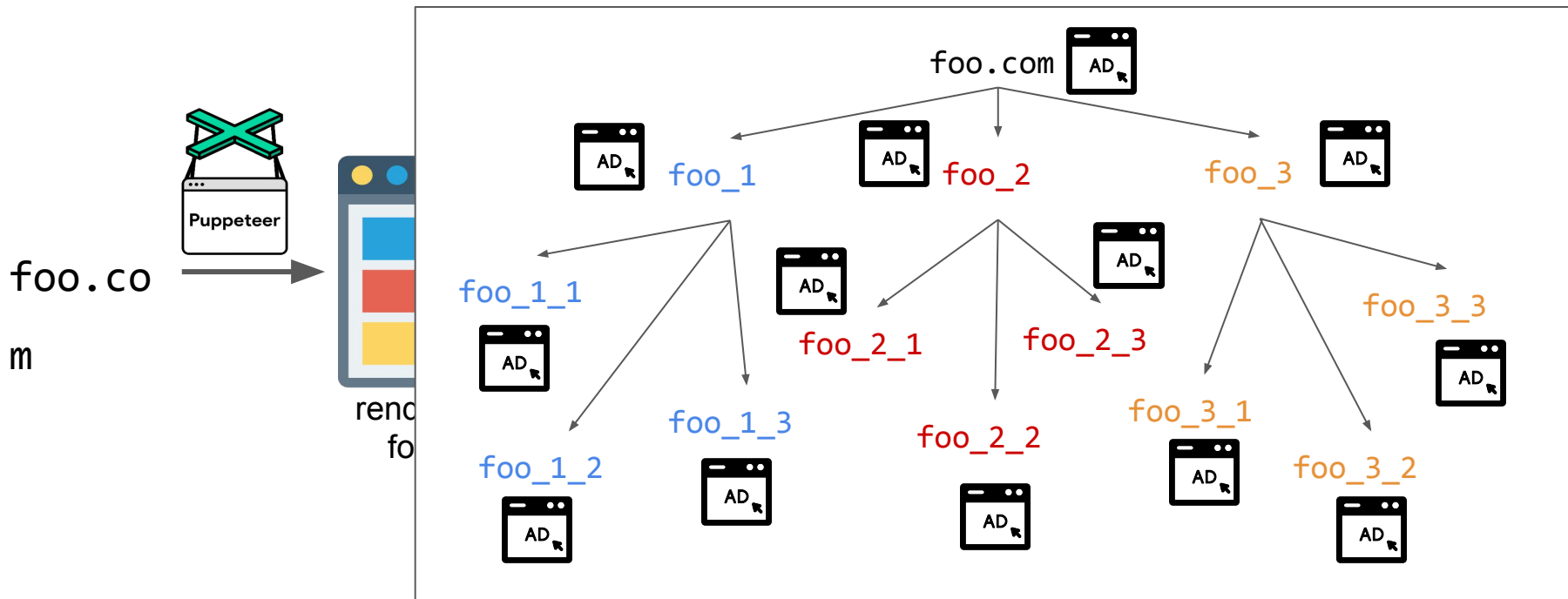
11

Research Problem

- Understanding the ecosystem of online

  ads on fake news sites

  - Identifying **third-party** mediators
  - Identifying **advertisers**
  - **Categorizing** ads (e.g., medical, health, etc.)
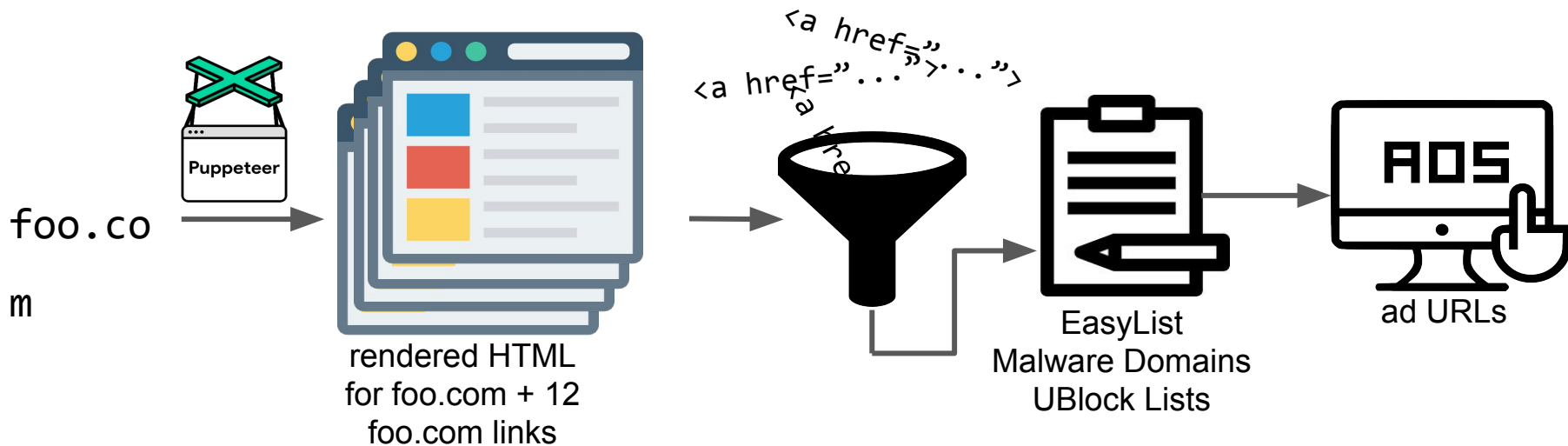
- Comparative analysis with popular benign

  news sites

Construction of Site Corpus

- Fake News
  - Zimdars' *False, Misleading, Clickbait-y and/or Satirical "News" Sources* List
    - Domain, "About Us", source, style, aesthetic, social media analysis

- "Real" (Benign) News
  - Alexa Top Sites (News category)
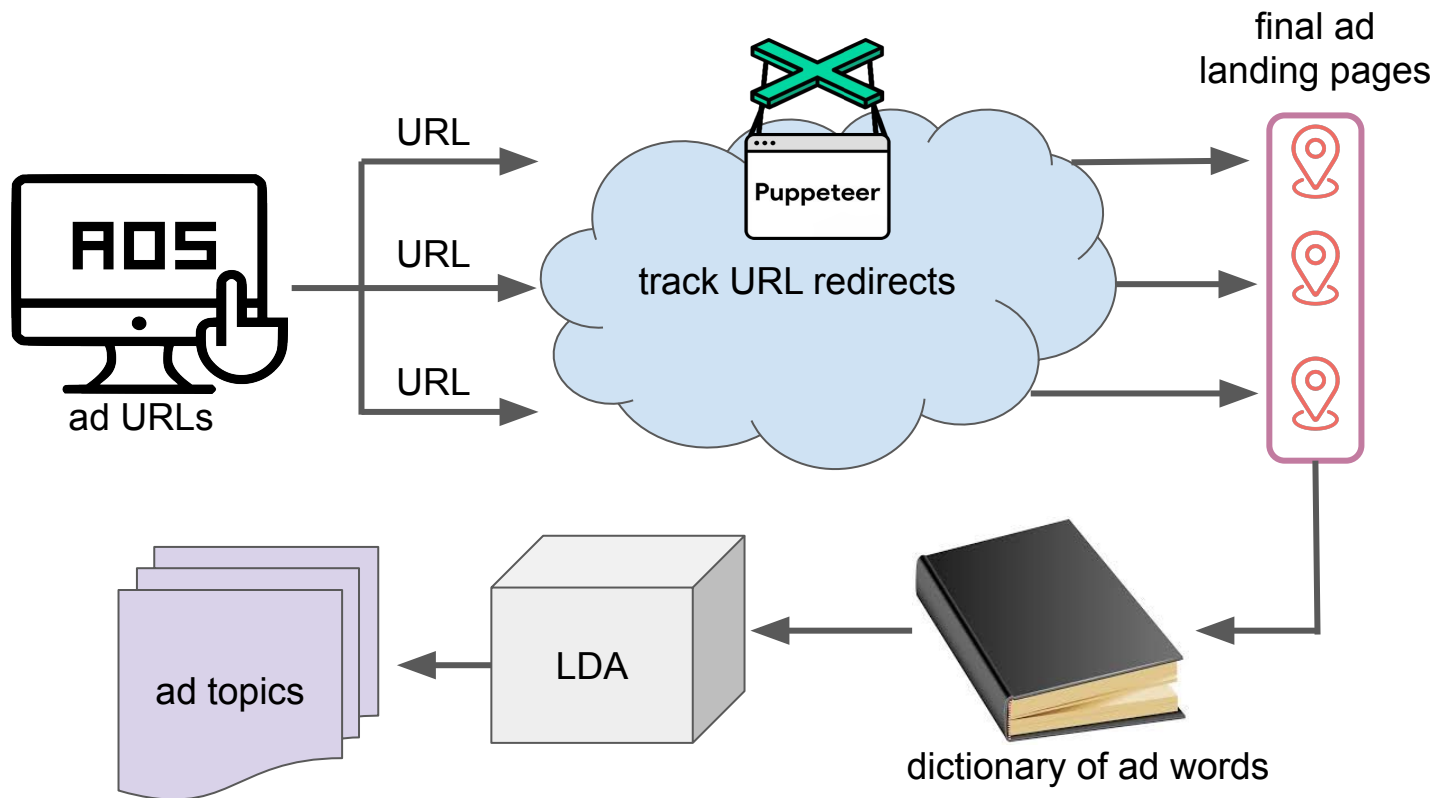
Methodology - Collecting Ads

Methodology - Collecting Ads

foo.com

rendered HTML
for foo.com + 12
foo.com links

<a href="..."><a href="...">...">

EasyList
Malware Domains
UBlock Lists

ADS

ad URLs

Methodology - Categorizing Ads

URL

URL

URL

ad URLs

Puppeteer

track URL redirects

final ad
landing pages

ad topics

LDA

dictionary of ad words
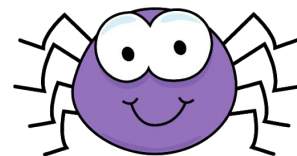
Methodology - Challenges & Lessons Learned

- Choosing the corpus of fake news sites

- Web crawling woes
  - Headless browsing detection
  - Dynamically loaded content (ads)
  - Asynchronicity in JS

- Categorization of websites
  - Amazon Web Information Service API

Limitations & Ethics

- Inability to account for fingerprinting

- Corpus size and robustness

- Gathering (consistent) corpus metadata

- Clicking on ads

18

Results - Third-Party Ad Servers

- Third-party ads were found on 246 fake news sites and 716 benign sites

Third-Party Ad Domain	Unique Fake News Sites
googleadservices.com	123
doubleclick.net	109
beverlyhillsmd.com	21
reduces-fingoloft.com	17
drmartypets.com	9
amazon-adsystem.com	7
iyfsearch.com	6
gundrymd.com	5
keywordblocks.com	5
google.com	4

…
(truncated)

Third-Party Ad Domain	Unique Popular Sites
doubleclick.net	596
googleadservices.com	443
gundrymd.com	56
criteo.com	54
constellationnutrition.com	46
mypowerlife.com	37
ultimatepetnutrition.com	33
activatedyou.com	33
unifyhealthlabs.com	29
lifestylejournal.com	24

…
(truncated)

Results - Site Ranking and Third-Party Ad Servers

Relationships With Ad Servers vs. Site Ranking
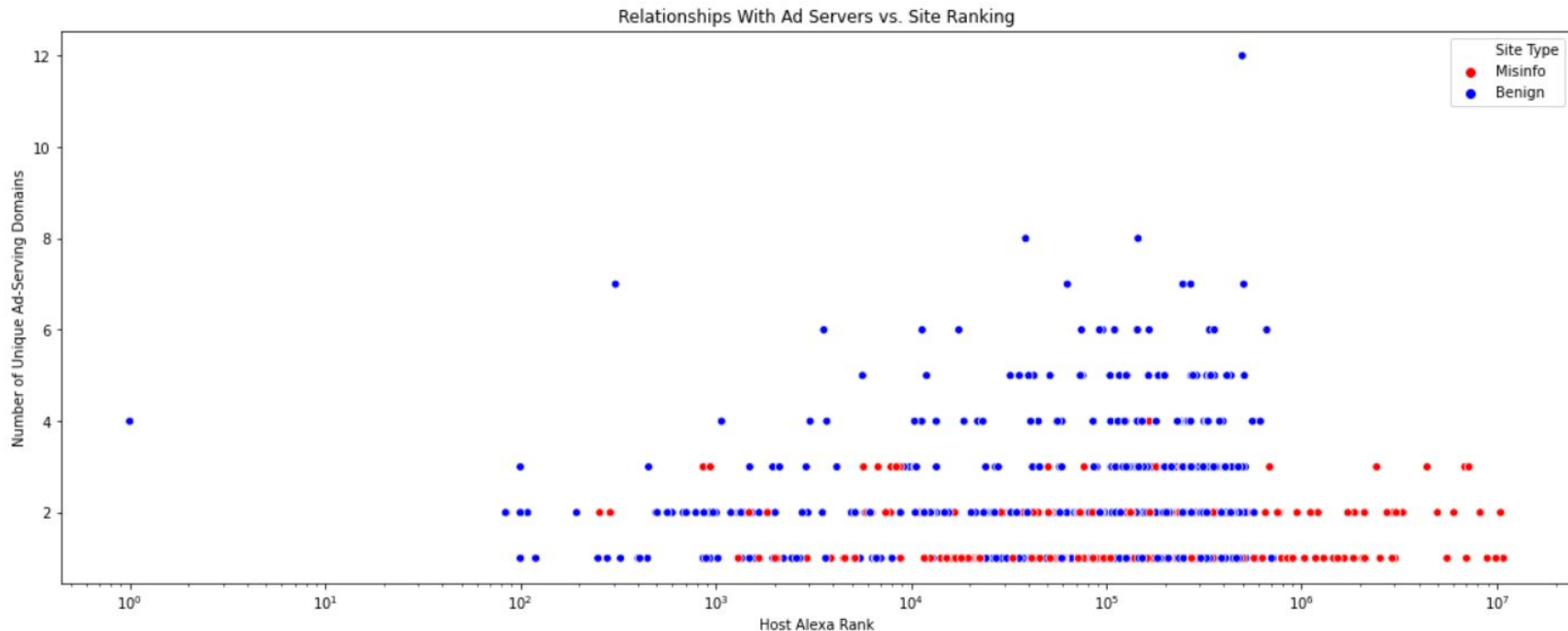
Results - Advertisers

Misinfo

Final Ad Domain	Host Count
vk457sprinklersettlement.com	552
bevhillsmd.com	378
healthydailybuzz.com	238
ifcj.org	182
graysonsettlement.com	168
sharecare.com	132
searchthese.net	130
bestsearches.net	110
amazon.com	100
techgadgetstrends.com	99

Benign

Final Ad Domain	Host Count
facebook.com	7569
globalindustrial.com	4356
usatoday.com	3844
doubleclick.net	3249
shutterstock.com	2916
thebarefootwriter.com	2809
cleanbodyenergy.com	2116
muscledefense.com	1296
insticator.com	1089
energeticallday.com	1089

Results - Site Ranking and Advertisers

Rankings of Host Sites Per Advertiser
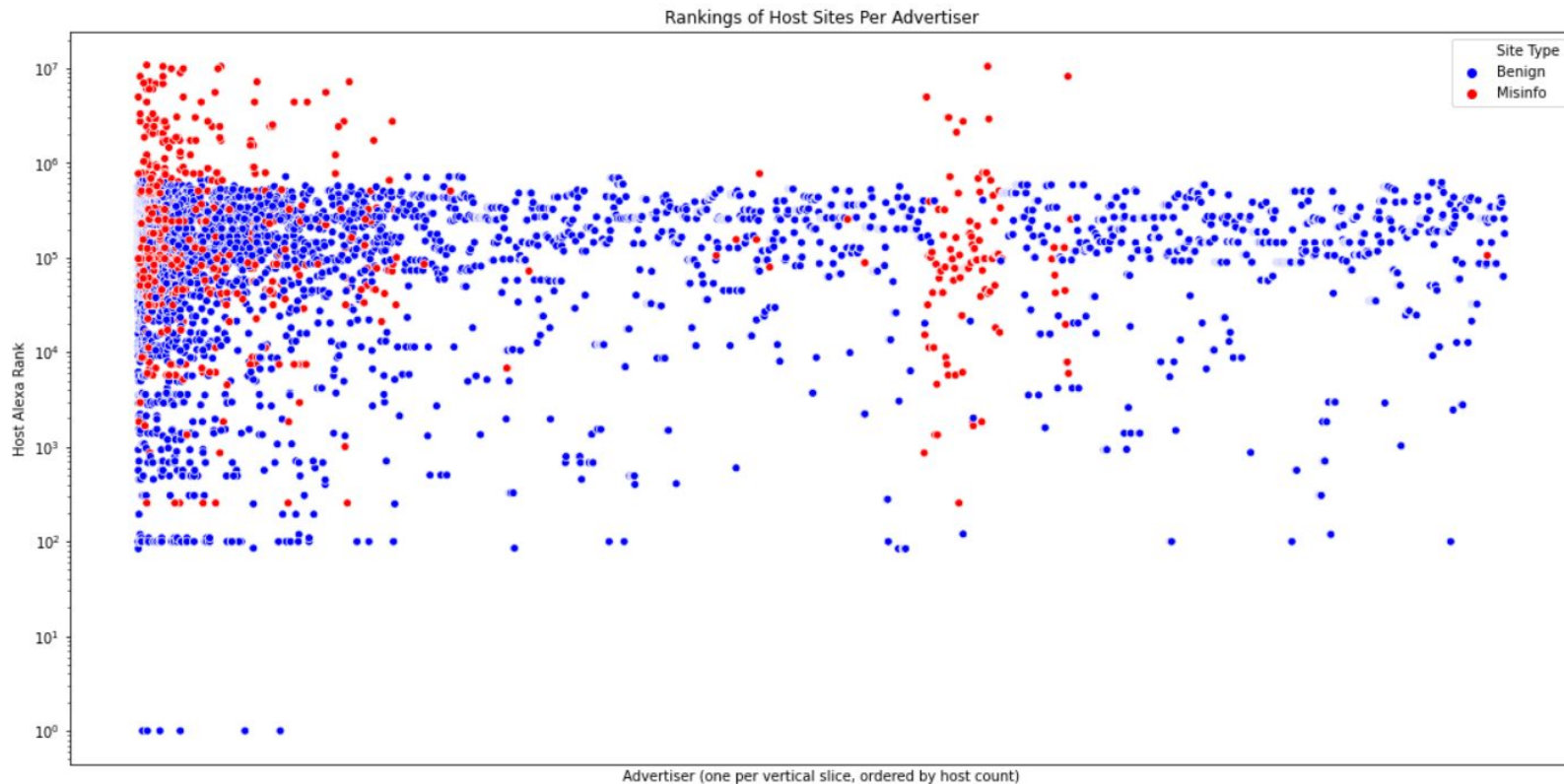
Results - Products Advertised

Advertisement Category	Unique Fake News Sites (N=246)	Unique Benign News Sites (N=716)
Technology	235 (96%)	403 (56%)
Medicine & Health	133 (57%)	383 (53%)
Coronavirus (proper subset of Medicine & Health)	36 (15%)	151 (21%)
Finance	84 (34%)	329 (46%)
Politics	24 (10%)	40 (6%)
Cannabis and CBD	23 (10%)	46 (6%)

Future Work

- Larger corpus

- Categorization of ads

- Longitudinal crawling

- Motivations of users navigating to such sites
  - Browser extension + user study with MTurk