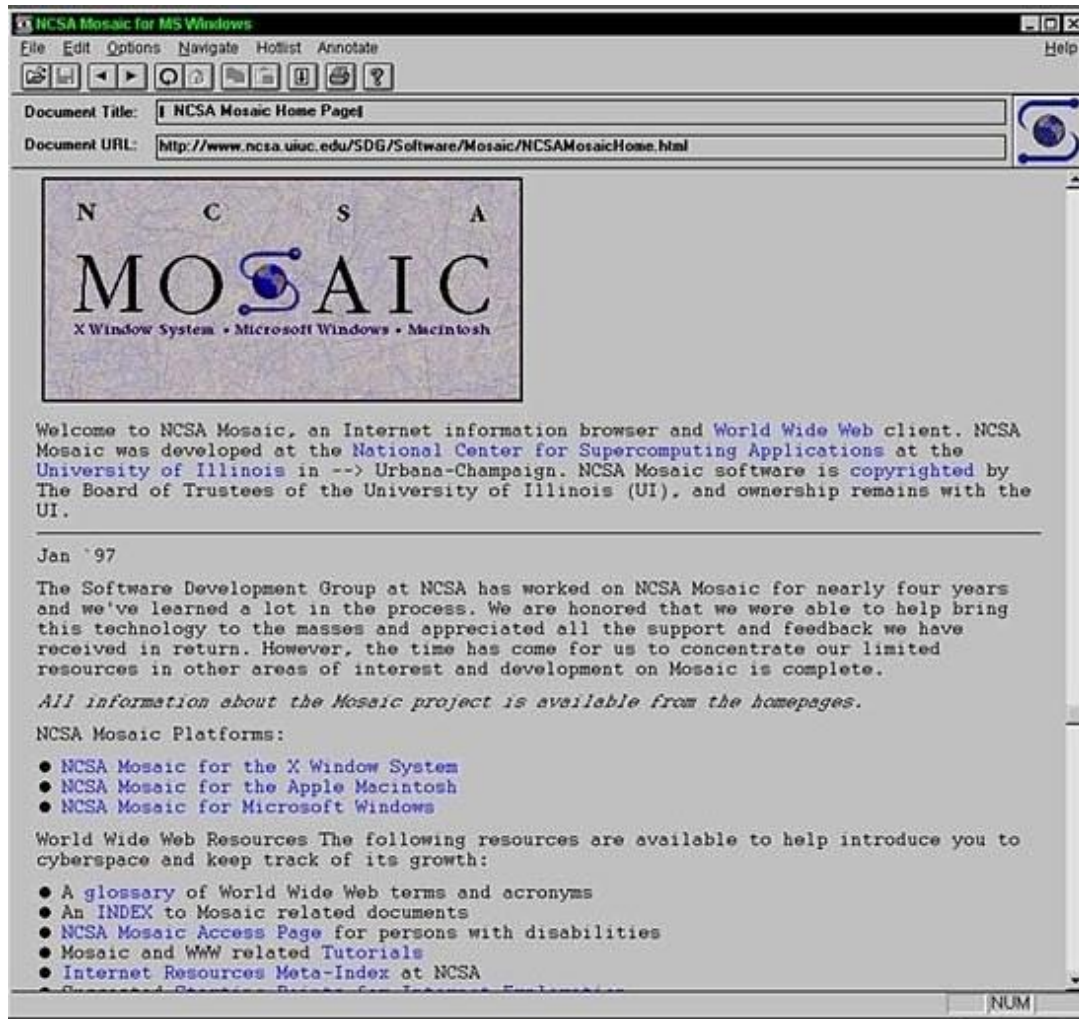# Beyond content analysis: Detecting targeted ads via distributed counting

Costas Iordanou (MPI); Nicolas Kourtellis (Telefonica Research); Juan Miguel Carrascosa (LSTech); Claudio Soriente (NEC Laboratories Europe); Ruben Cuevas (Universidad Carlos III de Madrid); **Nikolaos Laoutaris** (IMDEA Networks Institute)

First ever display ad (1994)

Some history

EVOLUTION OF DIGITAL MARKETING

1990 Prodigy, CompuServe

1991 America Online

1992 SMS Messaging

1994 First Online Display Ad

1995 Yahoo Search Engine

1996 Email Marketing

1997 AIS Media Launched, Mobile Payments, SEO

1998 SEM & Google

2000 PPC, AdWords, Google Mobile Advertising

2003 LinkedIn, WordPress, MySpace

2004 Mobile First, Yelp, Responsive Web Design, Facebook

2005 Google Analytics, SEO, YouTube

2006 Twitter, Marketing Automation

2007 iPhone, App Store, Geotargeting, Tumblr

2008 Facebook Ads

2009 Bing

2010 iPad, Instagram, Internet Surpasses Newspaper

2011 Google+, Pinterest, Snapchat

2012 Infographics, Visual Content Marketing

2013 iBeacon, Micro-location

2014 Mobile Surpasses Desktop Users, Omni Channel

2015 Wearables

2016+ Internet of Things, Mobile Overtakes Email & Social, Online Ad Spend Equivalent to TV Spend

1990   2000   2010   2020

AIS Media

⅓ of digital advertising is display (banners)

Half year 2018
($49.5 billion)

Half year 2019
($57.9 billion)

Half year 2018: Search 46%, Banner 32%, Video 14%, Other 8%

Half year 2019: Search 45%, Banner 31%, Video 16%, Other 8%

Search
Banner *
Video
Other **

https://www.iab.com/insights/internet-advertising-revenue-2019-half-year/
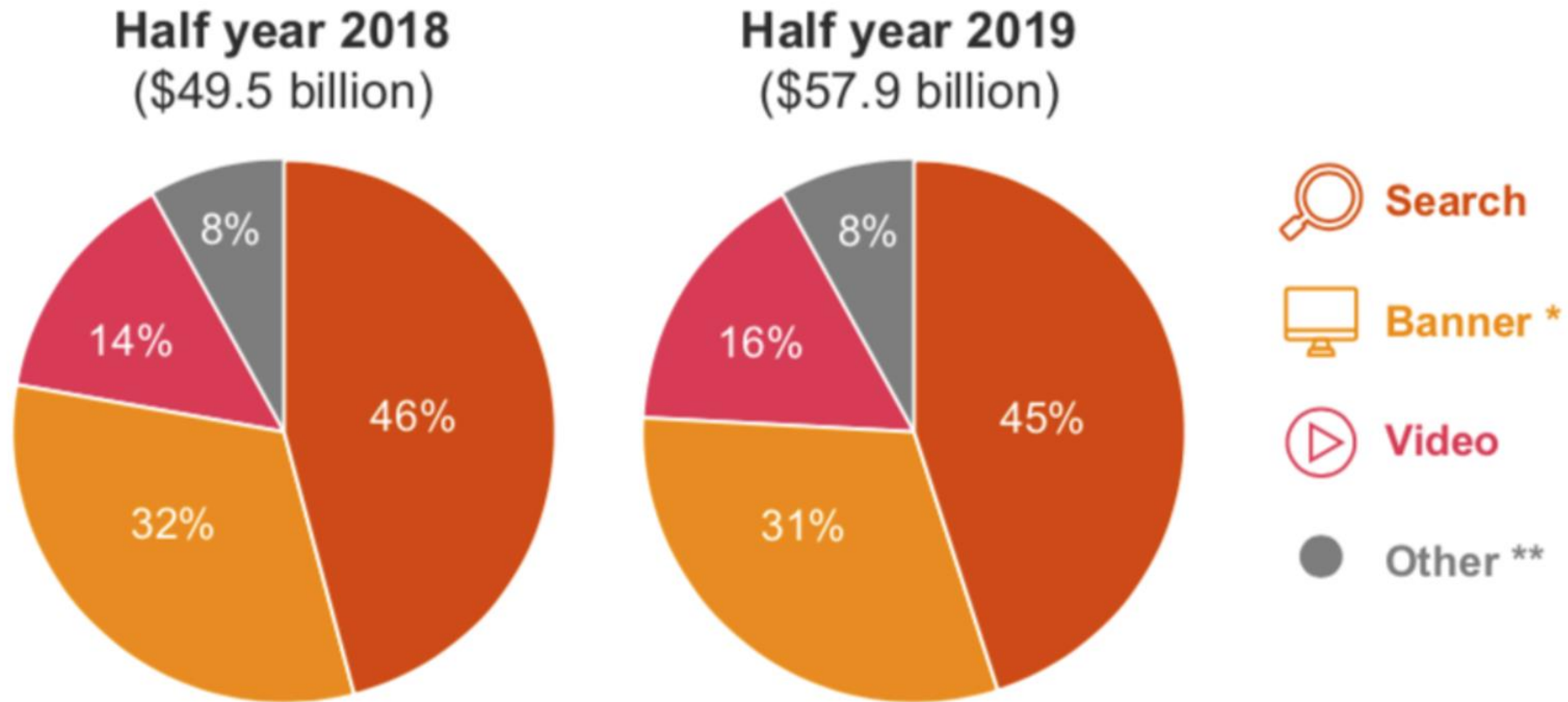
Programmatic RTB auctions in action

How about these ones?
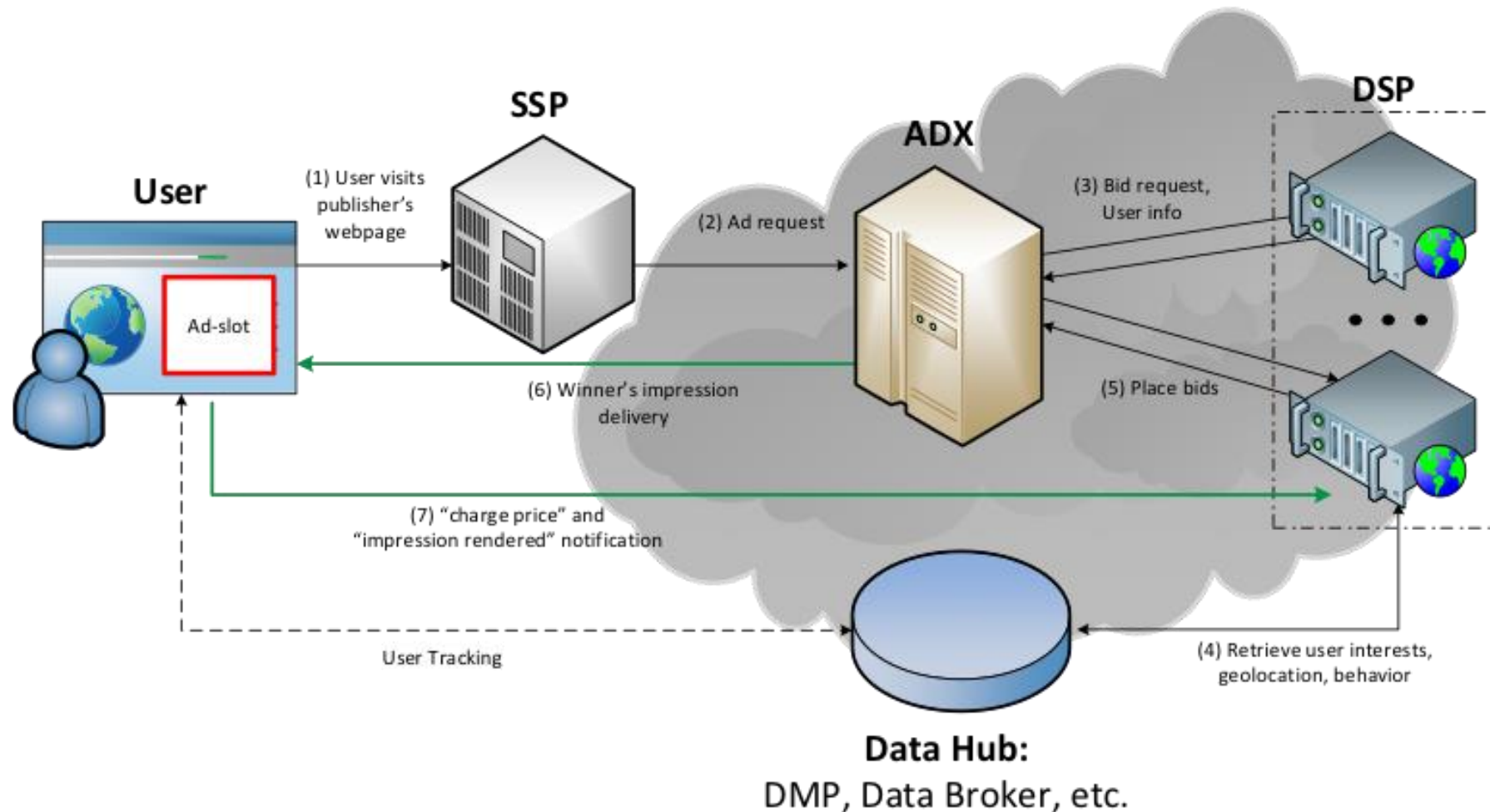
THESE ARE DISPLAY ADS

Is this ad targeted?
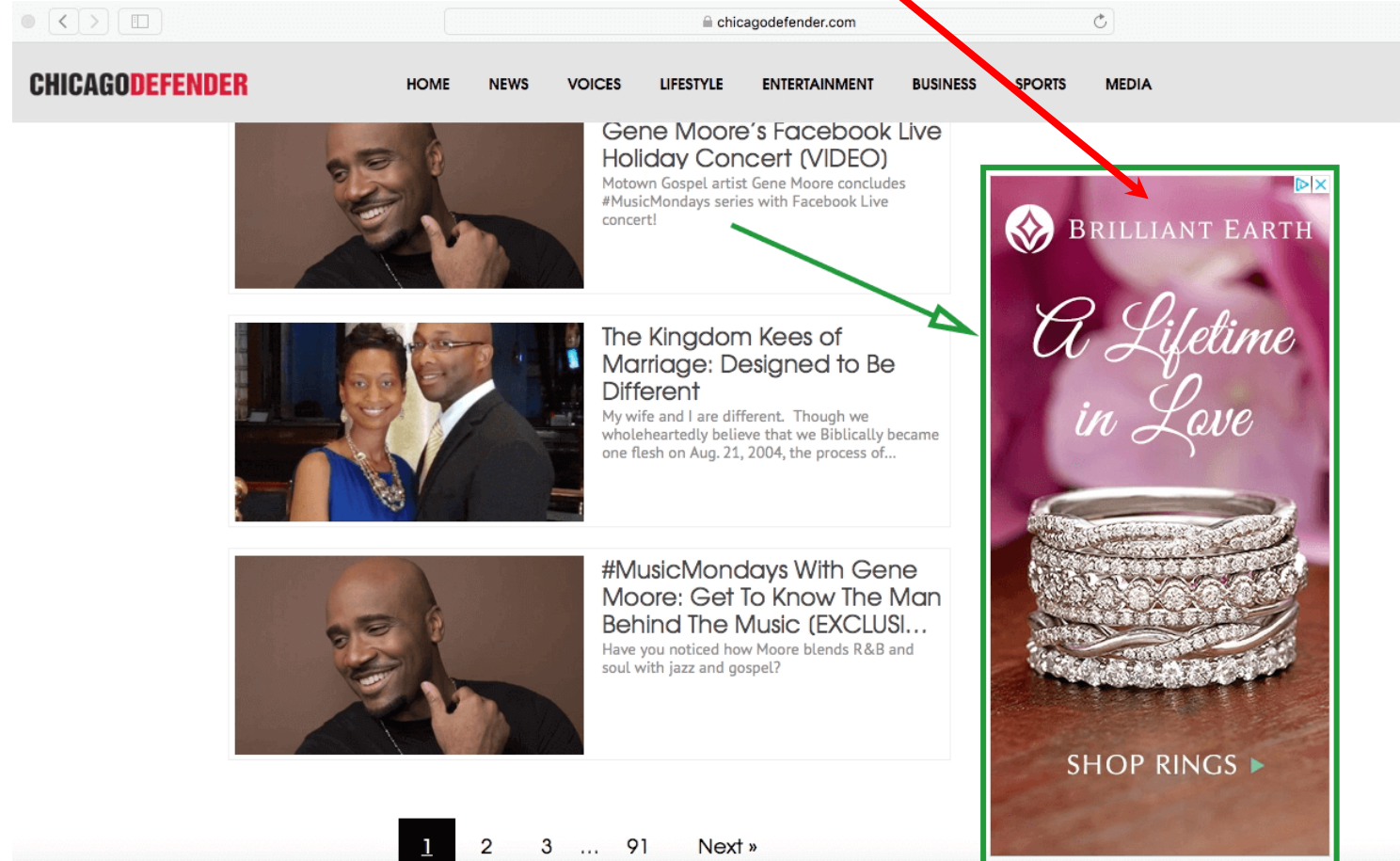
Why do we need to know if an ad has been targeted?

Enforcement of data protection laws & self-regulation initiatives:

- EU GDPR → no targeting on sensitive personal data
- FTC COPPA → no targeting on minors
- AdChoice, DNT → do they work?

Advertising analytics:

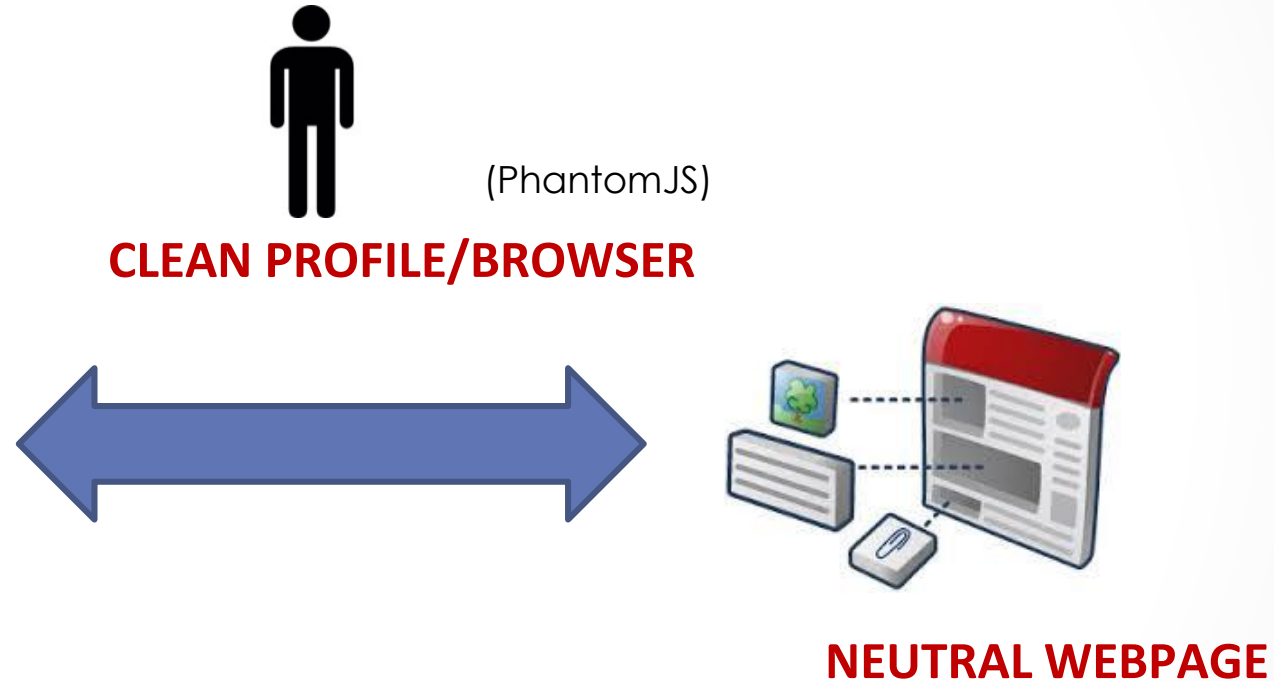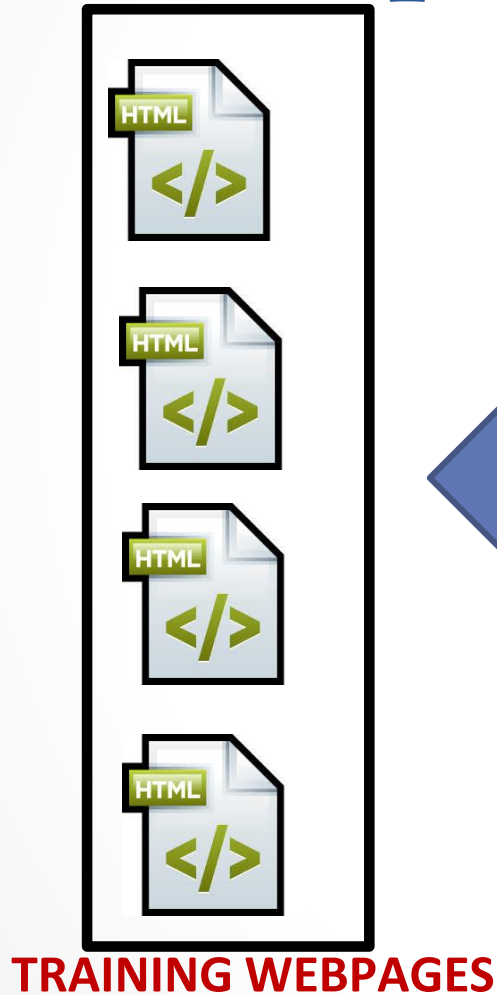- Who targets a particular audience?
- Which audiences are more targeted?

State of the art: Content-based detection

1. Label visited web-pages
2. Label received ads
3. Compute correlation

Passive → web-log analysis (Sunlight, Lecuyer et al., CCS'15)

Active → artificial "personas" (Carrascosa et al., CoNEXT'15)

Artificial persona based detection

OBSERVER

HTML </>

HTML </>

HTML </>

HTML </>

(PhantomJS)

CLEAN PROFILE/BROWSER

NEUTRAL WEBPAGE

TRAINING WEBPAGES

CoNEXT 2015

Limitations of content-based detection

- Complex
- Slow
- Not scalable
- Intrusive
- Cannot detect implicit targeting

New York

SUBSCRIBE | SIGN IN

VULTURE

2016 ELECTION | NOV. 22, 2016

Trump's Campaign Targeted His Supporters' Favorite TV Shows: *NCIS* and *The Walking Dead*

By Halle Kiefer

If you saw some very specific (and ostensibly, very compelling) pro-Trump commercials during your favorite programs this year, that's because Jared Kushner knows exactly what you like, both in terms of TV preferences and political priorities. In a new *Forbes* interview, the real-estate developer, husband of Ivanka Trump, and head of the Trump campaign's data operation explains in detail how he helped the president-elect utilize supporter data to create a targeted advertising strategy. For example, if you're a viewer who loves CBS and AMC, but hates the Affordable Care Act and the alleged threat of immigration, your viewing experience probably featured some Trump ads. As *Forbes* reports:

Photo: Gene Page/AMC

Sometimes it's all about changing view point

Earth vs. Sun
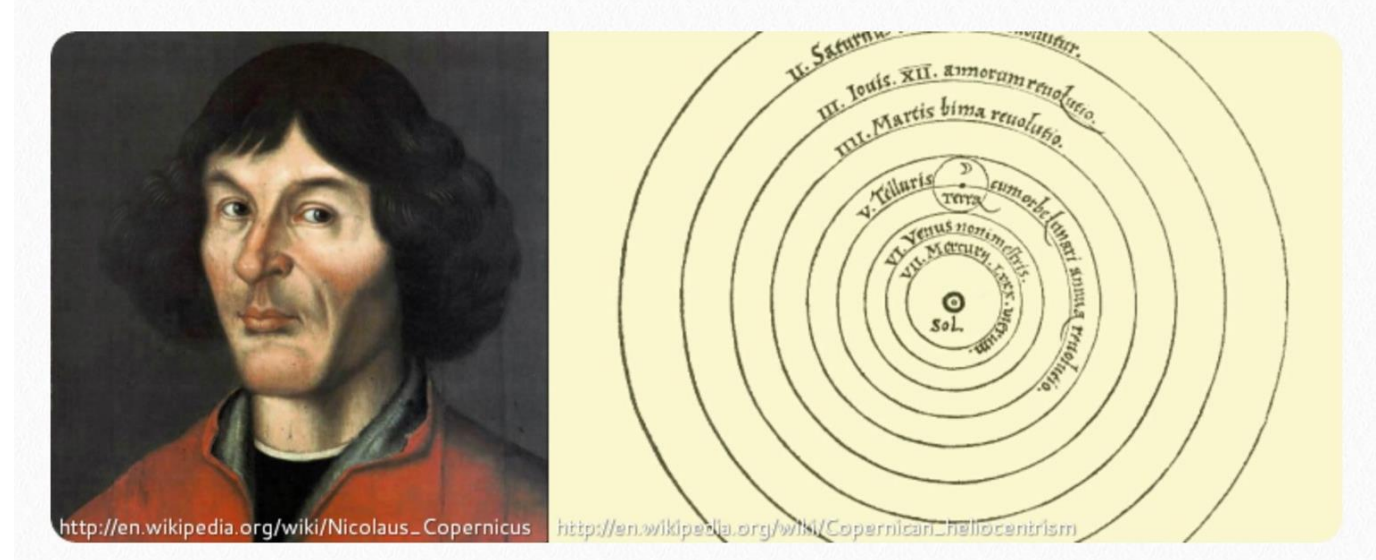
Content vs. Frequency
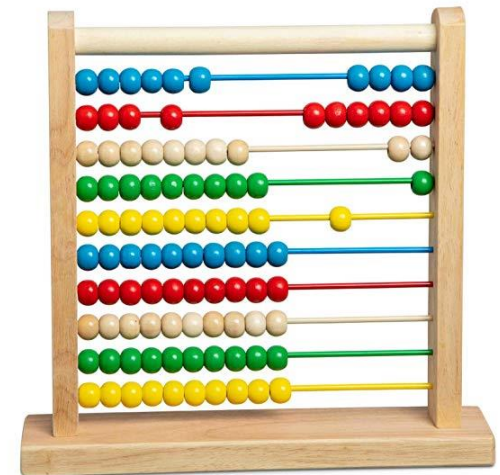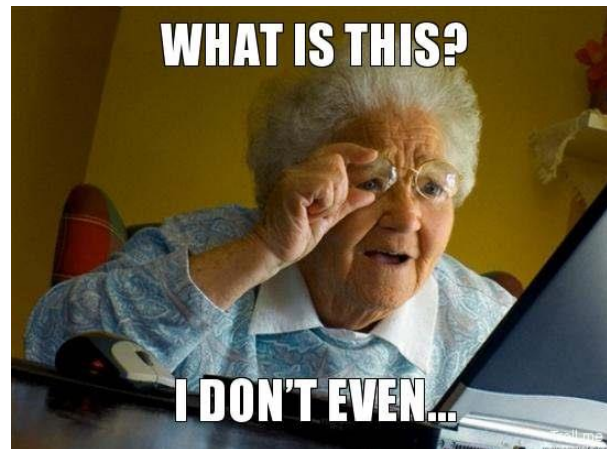
Targeted ads follow you around

- Detection via simple counting

- No need for content analysis

- No NLP, No ML

- No need to inject traffic

- Real-time

- No prob with indirect targeting

A simple algorithm

Algorithm 1 The count-based algorithm for ad α seen by user u

---

Require:

    Counters:

    $\#\text{Users}_\alpha$ $\triangleright$ Number of other users that observe ad α

    $\#\text{Domains}_{u,\alpha}$ $\triangleright$ Number of domains that user u observe ad α

    Thresholds:

    Users_{th} $\triangleright$ Users threshold based on all users

    $\text{Domains}_{\text{th},u}$ $\triangleright$ Domains threshold for a specific user u

1: **if** $\#\text{Users}_\alpha \leq \text{Users}_{\text{th}}$ **AND** $\#\text{Domains}_{u,\alpha} \geq \text{Domains}_{\text{th},u}$ **then**

2:        Targeted ad

3: **else**

4:        Non-targeted ad

Detects & annotates all rendered ads

Check your browsing history using the eyeWnder analysis tool: [Analysis Tool]
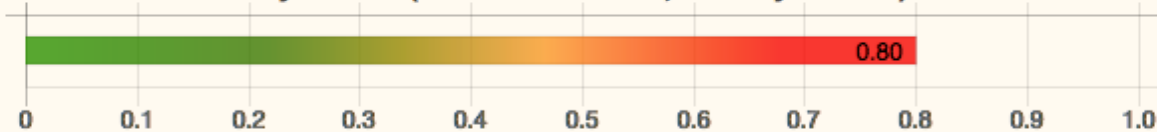
Selected Advertisement

User Feedback: Do you think it was a targeted ad? [▼]

Users Demographics Similarity

How similar are you to others that have seen this ad?

Seen by 3 users (0 = Not so Similar, 1 = Very Similar) Show more

0.80

| 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |

**Advertisement Information Table**

This ad takes you to:	https://subscribe.wired.com/subscribe/wired/103100?source=WIR_Footer_IntlTargeting_Apr16
This ad is about:	People & Society > Men's Interests Books & Literature > Magazines News > Technology News

[Analyze Advertisement]

[Get Intermediaries]

eyeWnder ©

Detecting targeted ads with eyeWnder

System architecture and information flow
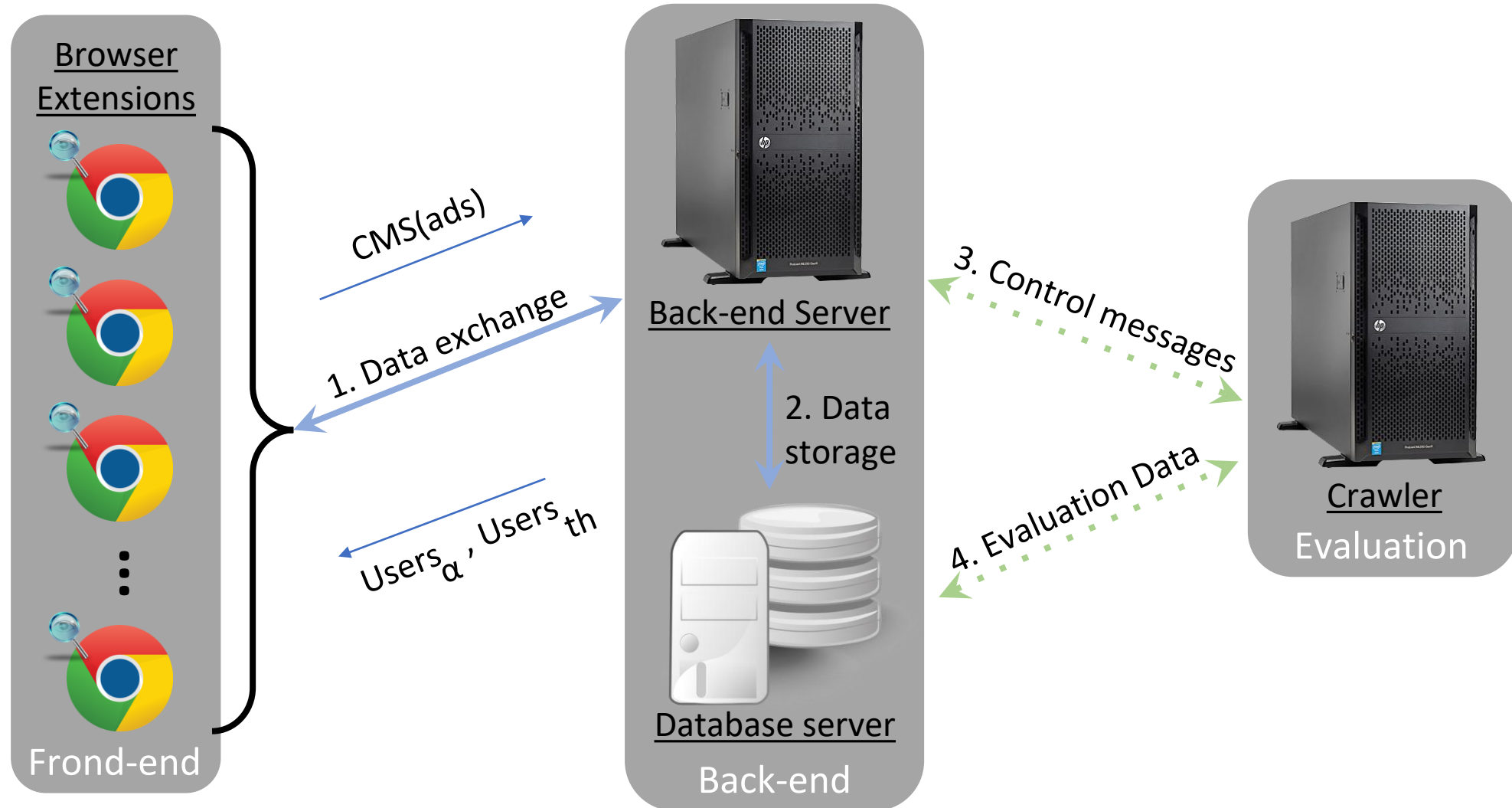
Detecting targeted ads with eyeWnder

Privacy preserving protocol
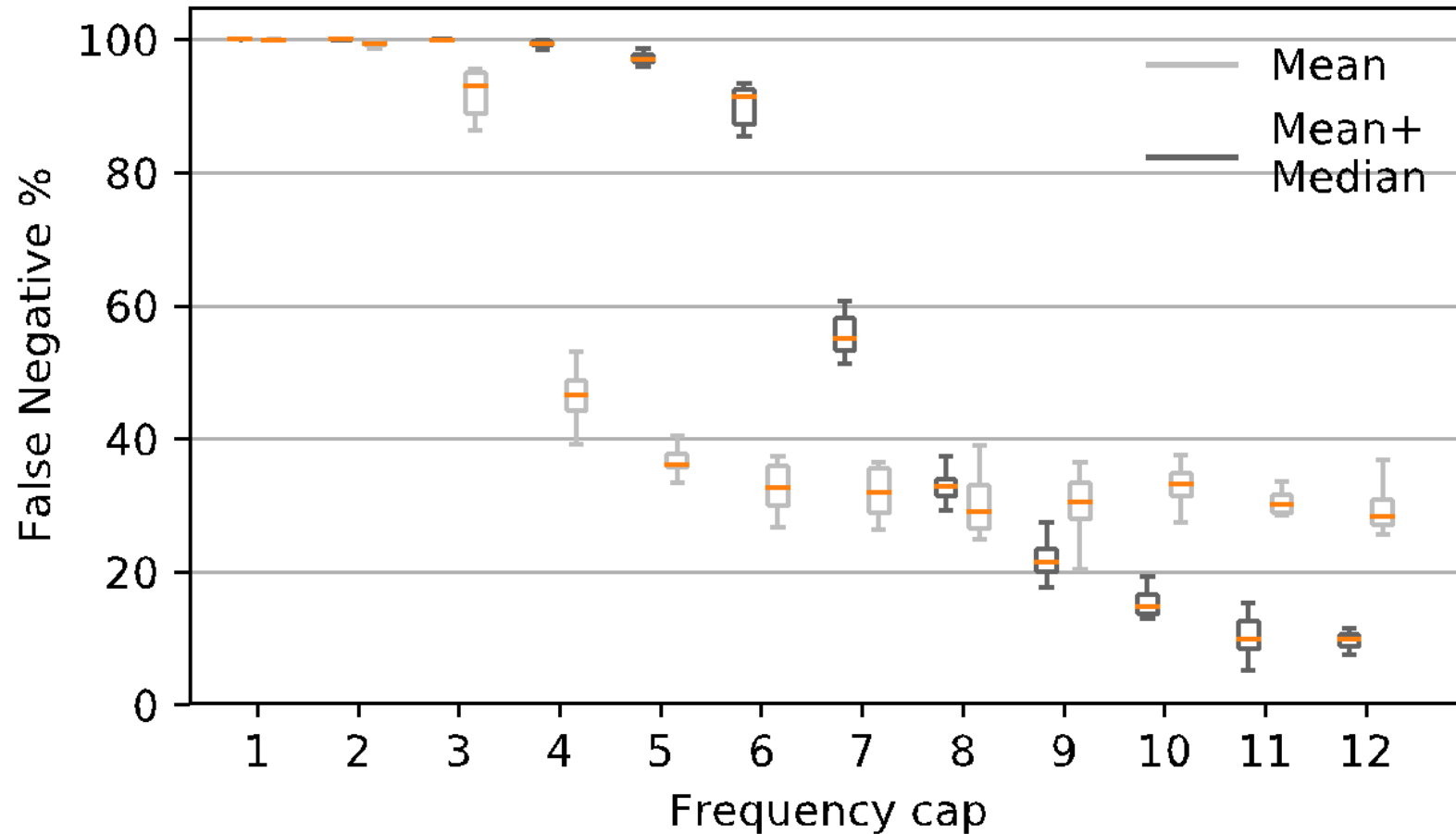
Impact of the privacy preserving protocol

The effect of the privacy preserving protocol on the computation of the #User distribution (number of users who saw a given ad α) and its threshold for three different weeks.

Simulation: <2% false positives, few false negatives

False Negatives % Vs. Frequency Cap using two different thresholds
(Mean, Mean + Median) for both variables (#Users$_\alpha$, #Domains$_{u,a}$)

(the most painful) Live validation (ever)

1000 users within 1 year

100 from CrowdFlower for 3 weeks

Users from all over the world
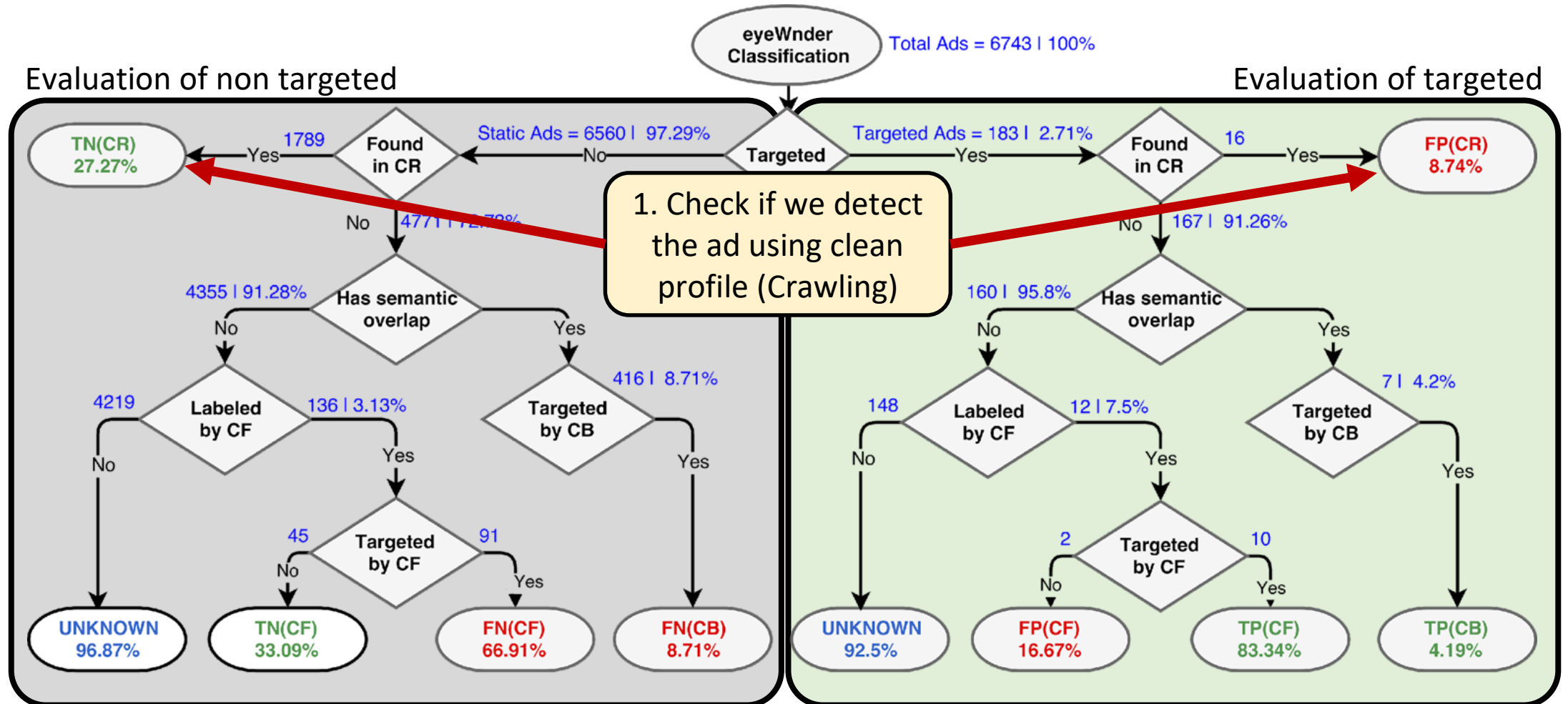
6743 ads within the 3 weeks control experiment

NO GROUND TRUTH!!!

Real time parallel crawling to detect static ads

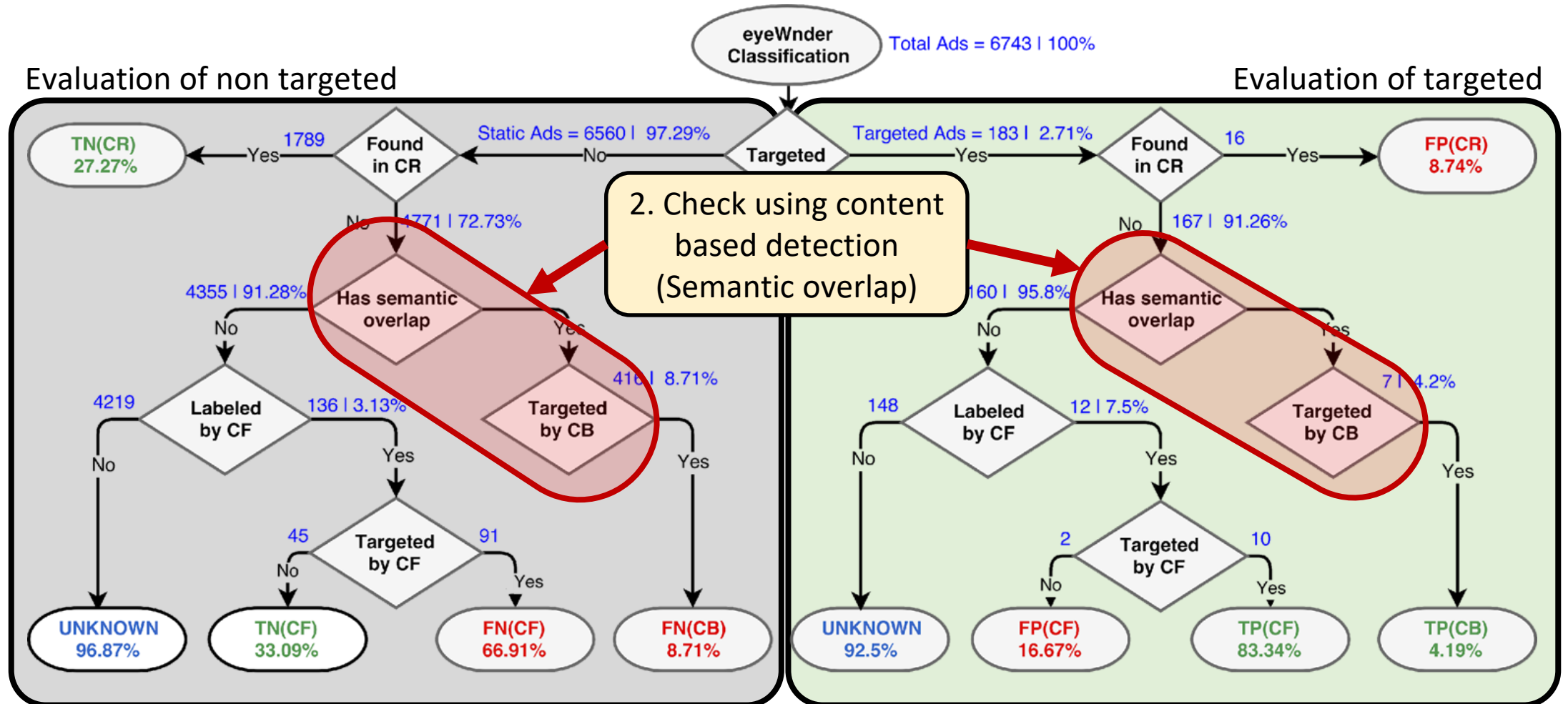Combination of manual and semi automatic inspection of targeted ads

Detecting targeted ads with eyeWnder
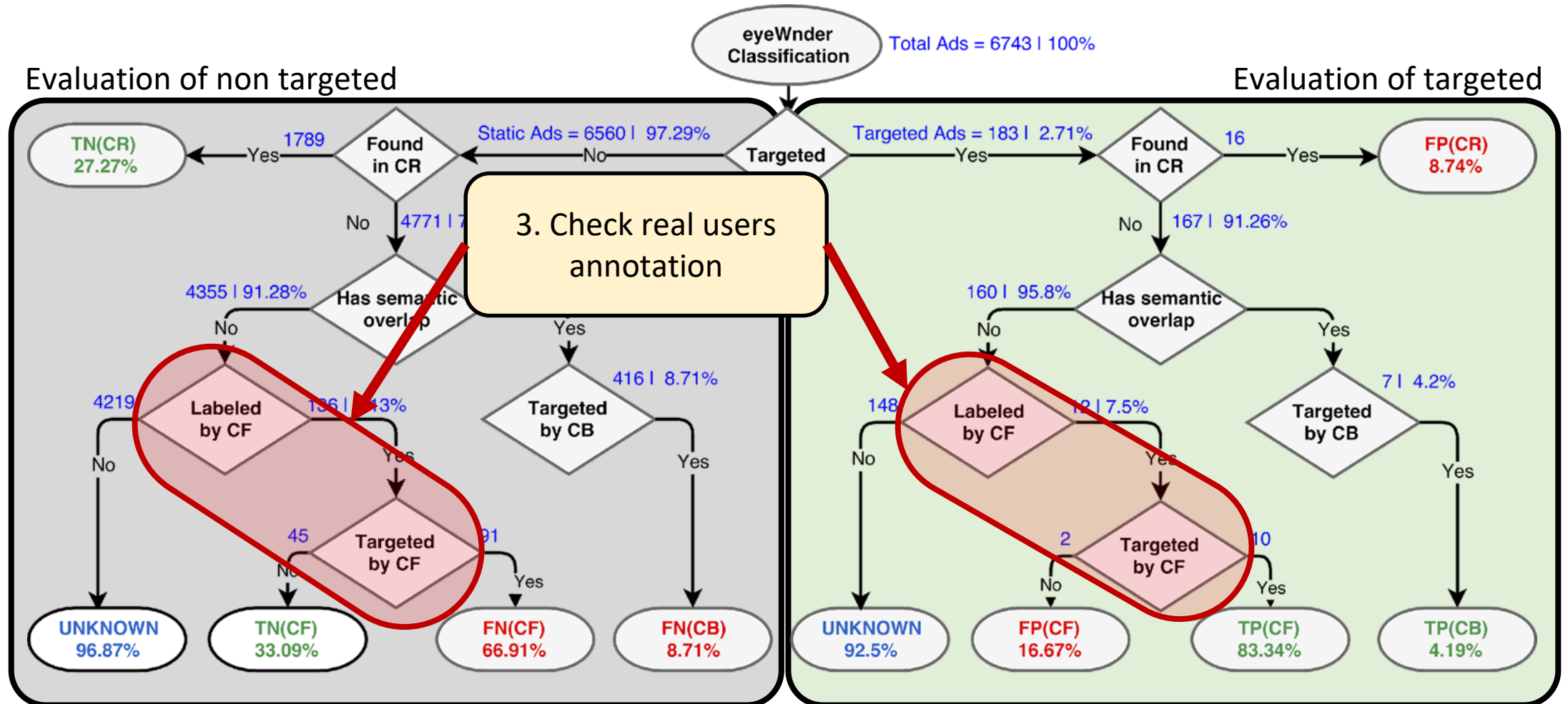
Evaluation results
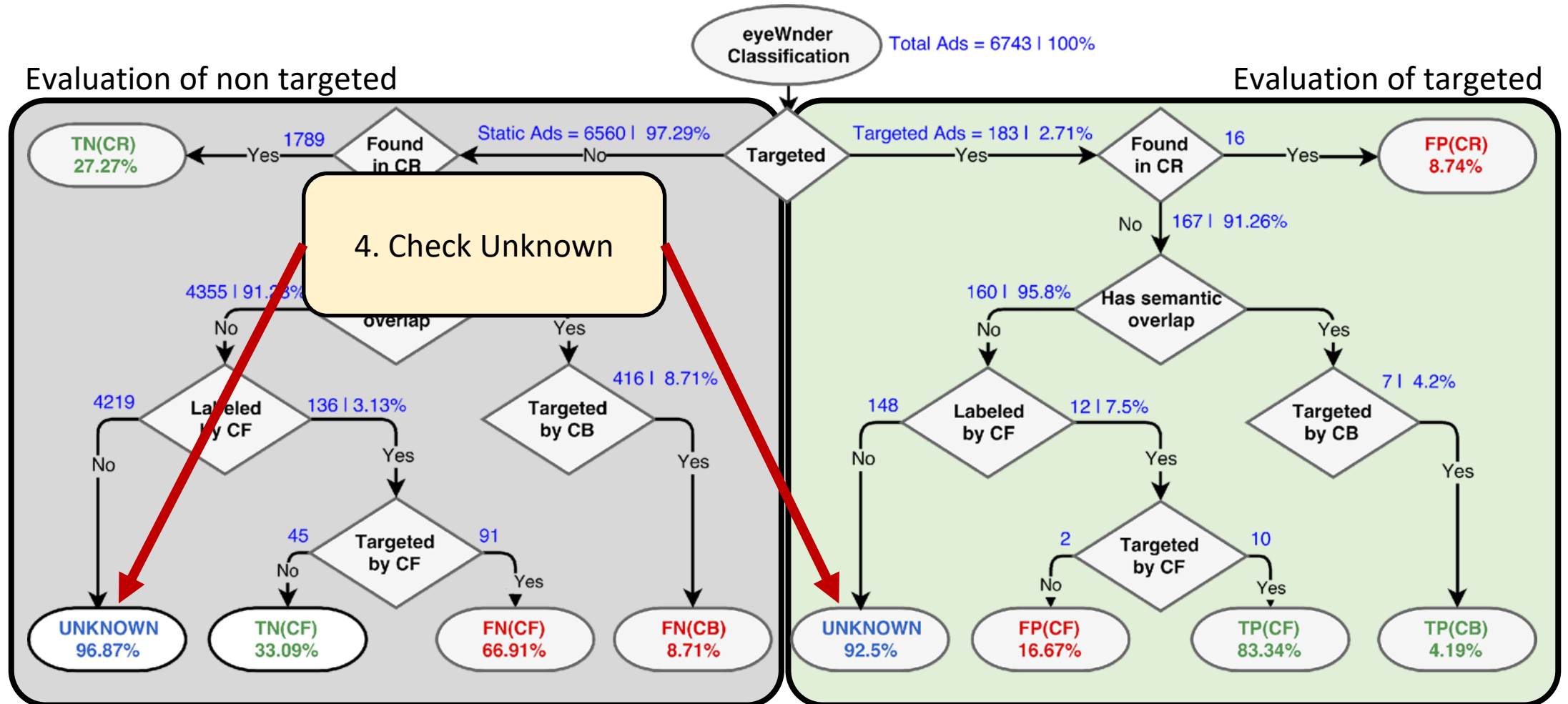
Detecting targeted ads with eyeWnder

Evaluation results

Detecting targeted ads with eyeWnder

Evaluation results

Evaluation of non targeted

Evaluation of targeted

eyeWnder Classification — Total Ads = 6743 | 100%

Static Ads = 6560 | 97.29% — Targeted — Targeted Ads = 183 | 2.71%

3. Check real users annotation

TN(CR) 27.27%

FP(CR) 8.74%

UNKNOWN 96.87%

TN(CF) 33.09%

FN(CF) 66.91%

FN(CB) 8.71%

UNKNOWN 92.5%

FP(CF) 16.67%
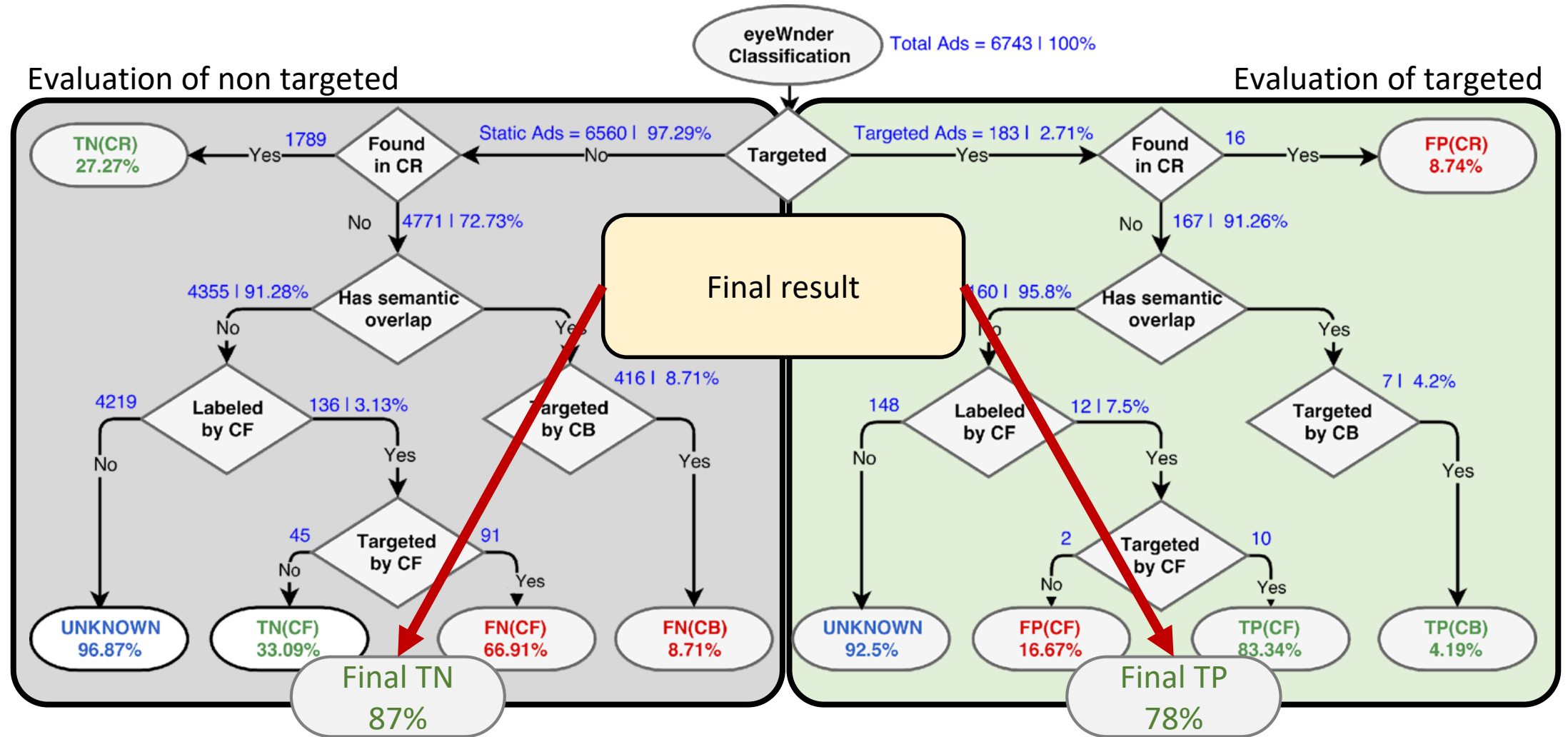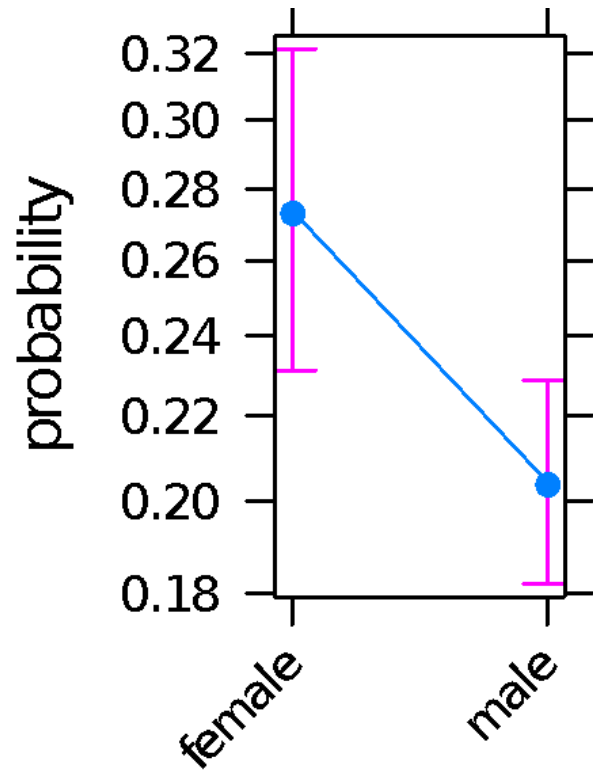
TP(CF) 83.34%

TP(CB) 4.19%

Detecting targeted ads with eyeWnder
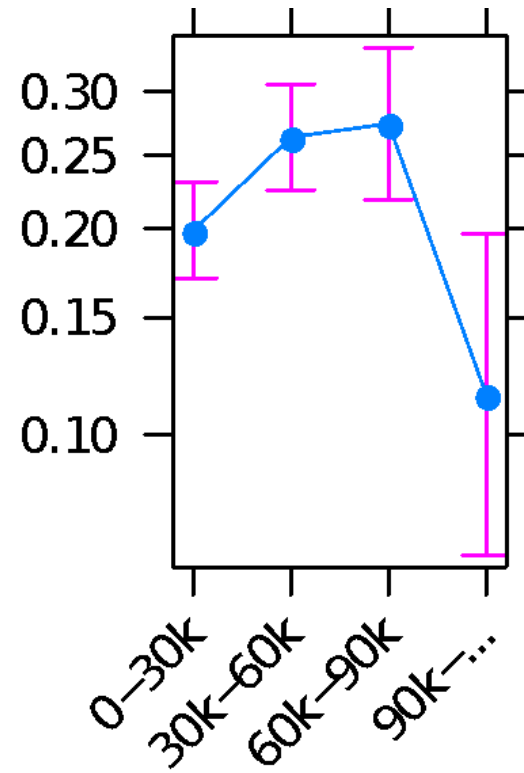
Evaluation results

Detecting targeted ads with eyeWnder

Evaluation results

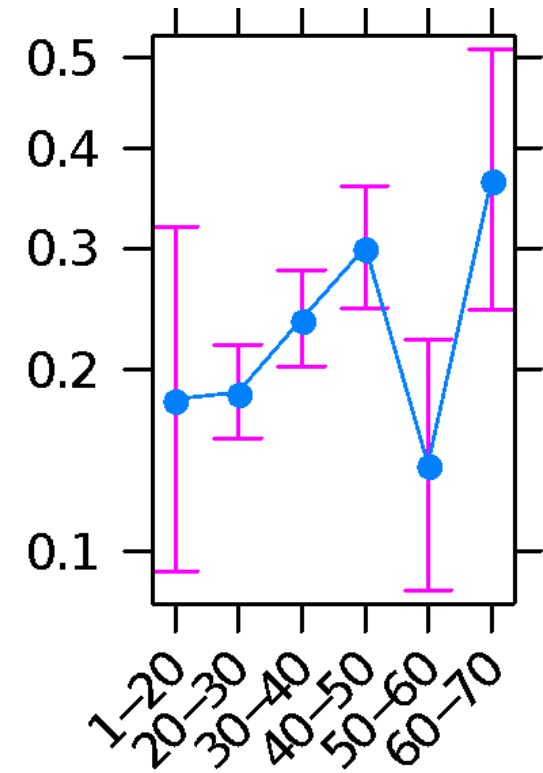# Socio-economic analysis results

Predicted probability for a targeted advertisement to be delivered to a user, vs. three independent variables with statistically significant levels.

Conclusions

Crowdsourcing makes targeted detection easy

The price for this is the need for some privacy preserving analytics

We have presented a first simple algorithm for detection

How much can it be improved?

More and larger experiment?