

Measuring Online Behavioural Advertising

A tale of
Transparency
&
Human-Centric Economics

Nikolaos Laoutaris

Research Professor
IMDEA Networks Institute

I am not a secretive person



I am not a privacy nut



Nikolaos Laoutaris

Nikolaos Laoutaris

Timeline About Friends 449 Photos More

1 Pending Item

Intro

A humble vaudevillian veteran cast vicariously as both victim and villain by the vicissitudes of Fate

- Senior Researcher at Telefonica
- Former Researcher at Telefonica
- Former Postdoc Fellow at Harvard University
- Former Marie Curie Postdoc Fellow at Boston University
- Studied Computer Science at National and Kapodistrian University of Athens
- Studied Computer Science at National and Kapodistrian University of Athens
- Went to Moschato

Add Featured Photos

Status Photo / Video Life Event


What's on your mind?

Friends Post

Nikolaos Laoutaris

September 3 at 1:04pm · 🌐

Exploding cigar! Really? Did they buy it from Coyote Wile?



10 Ways the CIA Tried to Kill Castro

There are only so many different ways you can ambush someone with a sharpshooter, so some of the ways the CIA plotted to kill Castro were pretty wild.

real name

real Muhammad Ali pose

silly comment

About Me



I am a research professor at IMDEA Networks Institute in Madrid. Prior to that I was director of data science at Eurecat and chief scientist of the Data Transparency Lab which I co-founded in 2014 during my 10 year tenure as a researcher and senior researcher of Telefonica Research in Barcelona. Before Telefonica, I was a postdoc fellow at Harvard University and Marie Curie postdoc fellow at Boston University. I got my PhD in computer science from the University of Athens in 2004. My interests include: privacy/transparency/data protection, economics of networks and information, intelligent transportation, distributed systems, protocols, and network measurements.

lots of opinions



RECENT POSTS

- [Why online services should pay You for Your data? The arguments for a Human-Centric Data Economy](#)
- [Networking Research: Present, Future and Beyond](#)
- [PhD position available on the Economics of Personal Data](#)
- [Data Transparency: Concerns and Prospects](#)
- [Myth-busting: Most tracking flows on European citizens DO NOT terminate outside EU28 GDPR borders](#)
- [The three types of research papers and how I learned to recognise them](#)
- [There I said it: The Net Neutrality “debate” is the Climate Change “debate” of the Internet](#)
- [DTL Award Grants’17 announced!](#)
- [A brief farewell after 10 years](#)
- [Online advertising, data protection, and privacy concerns of users, industry, and regulators \(Video\)](#)

RECENT COMMENTS

- Vasileios on [There I said it: The Net Neutrality “debate” is the Climate Change “debate” of the Internet](#)

acm advertising applications arxiv bulk caching capacity
computer conex content data debate
distributed energy ieee ifip imc impact infocom
internet mobile netecon networks online optimal overlay
packet parallel performance personal playout price privacy receivers
replacement replication research routing schedulers search selfish sharing sigcomm social
streaming systems transactions transfers transparency video

in 2012 I gave up on all my previous research

to work exclusively on privacy



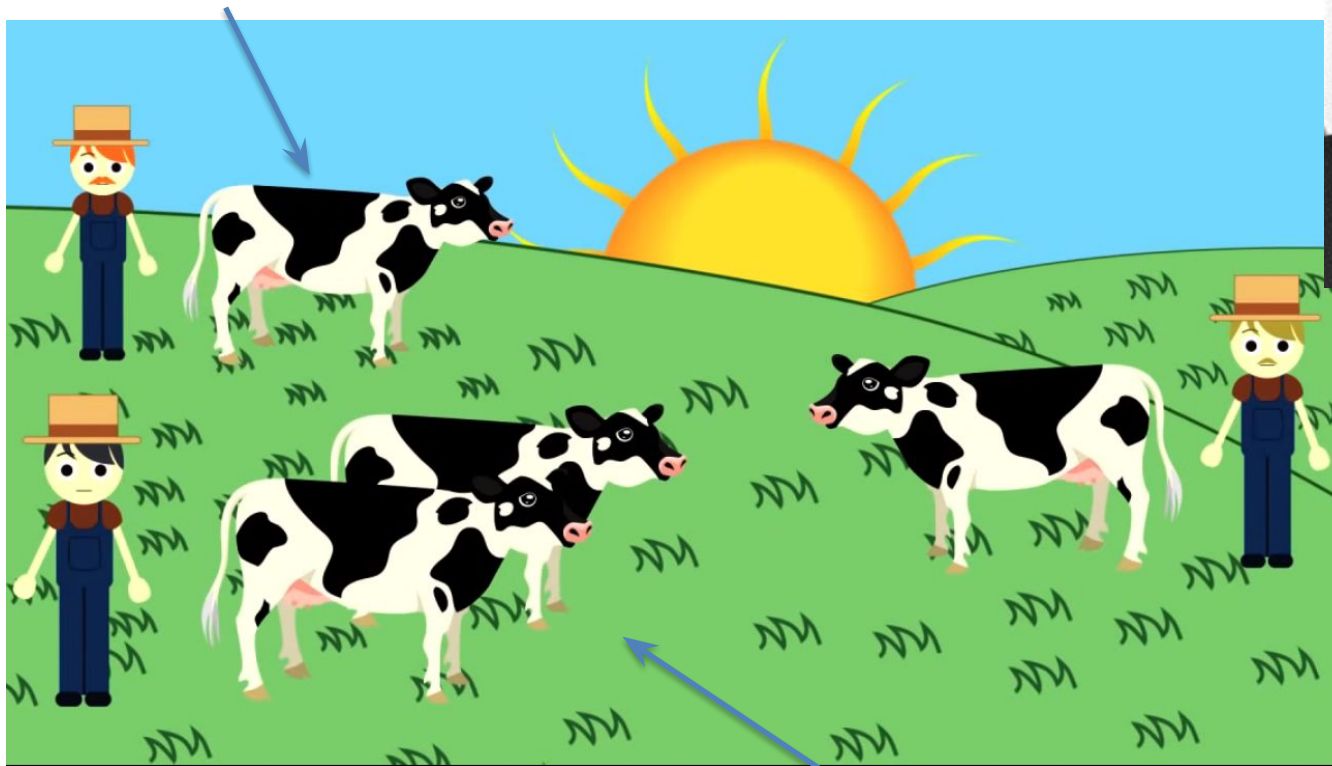
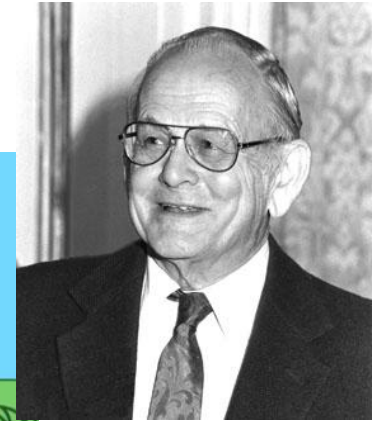
the web economy could collapse



due to Tragedy of the commons around privacy

Garrett Hardin, 1968

Internet company in
Web Economy ... crossing privacy **red lines**



The “commons”: consumer trust on the web and it’s business models

Big Idea #1 - Obvious in retrospect

The importance of Transparency (Software)



“Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.”

—U.S. Supreme Court Justice Louis D. Brandeis, “What Publicity Can Do,” *Harper’s Weekly*, December 20, 1913

A first of its kind Transparency Tool



\$heriff

Detecting Price
Discrimination

1. Select price

Availability: In Stock

Price: ~~\$200.50~~ \$189.00

Ex Tax: \$90.00

Price in reward points: 400

10 or more \$105.40

20 or more \$92.48

2. Check it

189.00

Check it

3. Examine differences

S, Safari, Spain	\$189.00
Firefox, Spain	\$189.00
m, Liège	\$165.99
São Paulo	\$189.00
d, Tampere	\$189.00
any, Berlin	\$201.50



Jakub Mikians
UPC (now Amazon)



Kostas Iordanou
Telefonica-UC3M

You might also like

- Accessories > Spare Batteries > For Canon
- Others > Cleaning Kits
- Memory > SD / SDHC / SDXC > 32GB
- Lenses > Sigma > Standard Fixed

Shop > DSLR > Canon EOS > EOS 7D Mark II

Canon EOS 7D Mark II Body

Fuel Your Creative Passion

- 20.2 MP APS-C CMOS sensor
- 3.0 inch (1040k) Clear View II LCD monitor

€ 1,229.01

In Stock (Ships within 24 hours)

Add to Basket



H1: 25600, H2:

F
S

All Prices Results

Results Details

Variant	Converted Value	Original Text
Canada, Ontario	€ 1409.60	C\$ 2,049.00
Czech Republic, Praha	€ 1395.52	US\$ 1,558.80
France, Champagne-ardenne, Troyes	€ 1249.84	€ 1,249.84
Ireland, Westmeath	€ 1249.84	€ 1,249.84
Portugal, Praga	€ 1249.84	€ 1,249.84
You	€ 1229.01	€ 1,229.01
Windows 7, Chrome, Spain	€ 1229.01	€ 1,229.01
Mac OS, Safari, Spain	€ 1229.01	€ 1,229.01
Linux, Firefox, Spain	€ 1229.01	€ 1,229.01
United States, Washington	€ 1162.94	US\$ 1,299.00
Poland, Warsaw	€ 1430.41	US\$ 1,597.77
Singapore, Singapore	€ 1250.16	US\$ 1,396.43
Australia, Clayton	€ 1175.79*	AU\$ 1,739.00
Brazil	€ 1453.67	US\$ 1,623.75

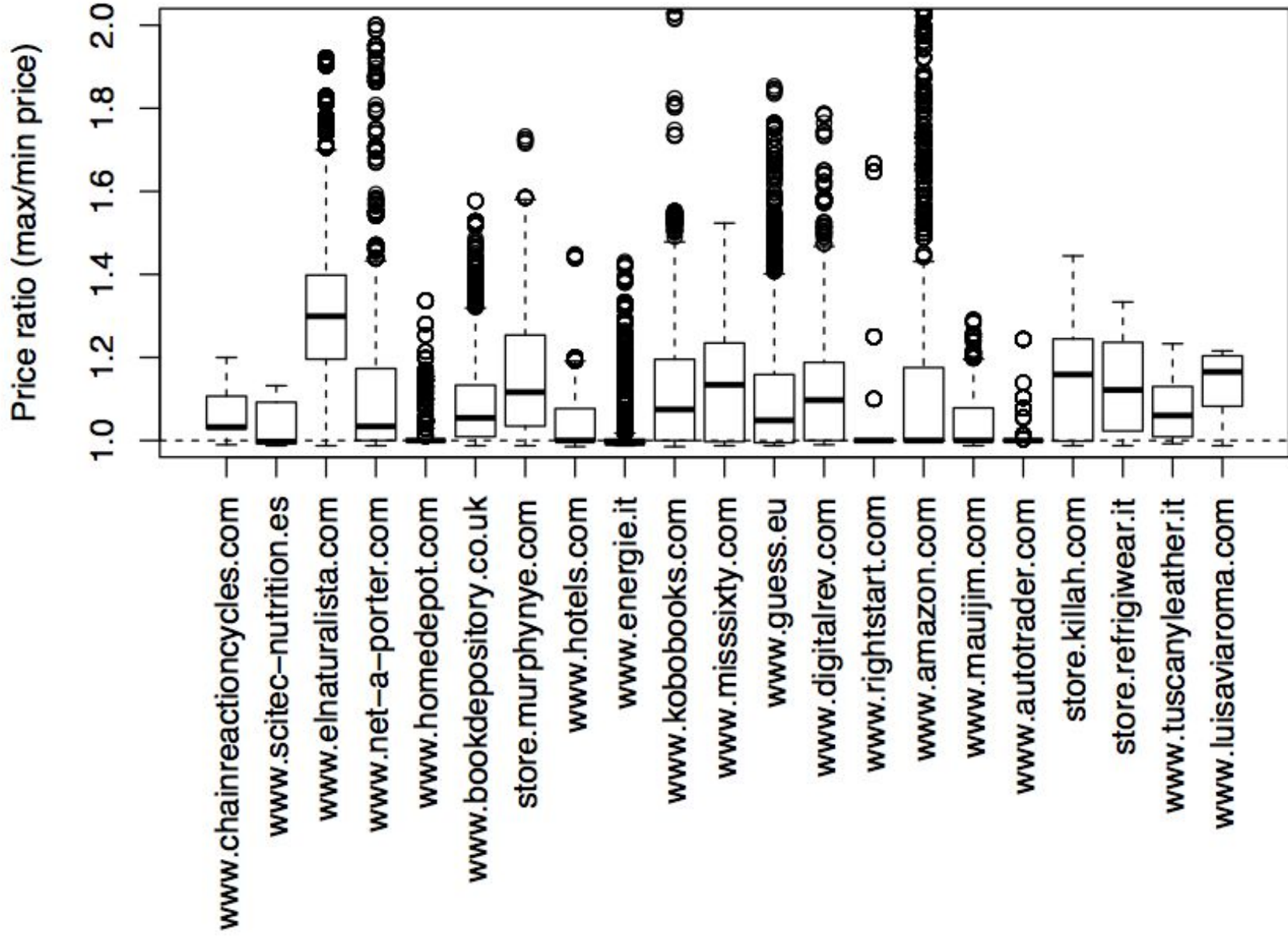
Results from local users

Source ID	Converted Value	Original Text
Local User 0	€ 1229.01	€ 1,229,01
Local User 1	€ 1229.01	€ 1,229,01
Local User 2	€ 1229.01	€ 1,229.01
Local User 3	€ 1229.01	€ 1,229.01
Local User 4	€ 1229.01	€ 1,229,01

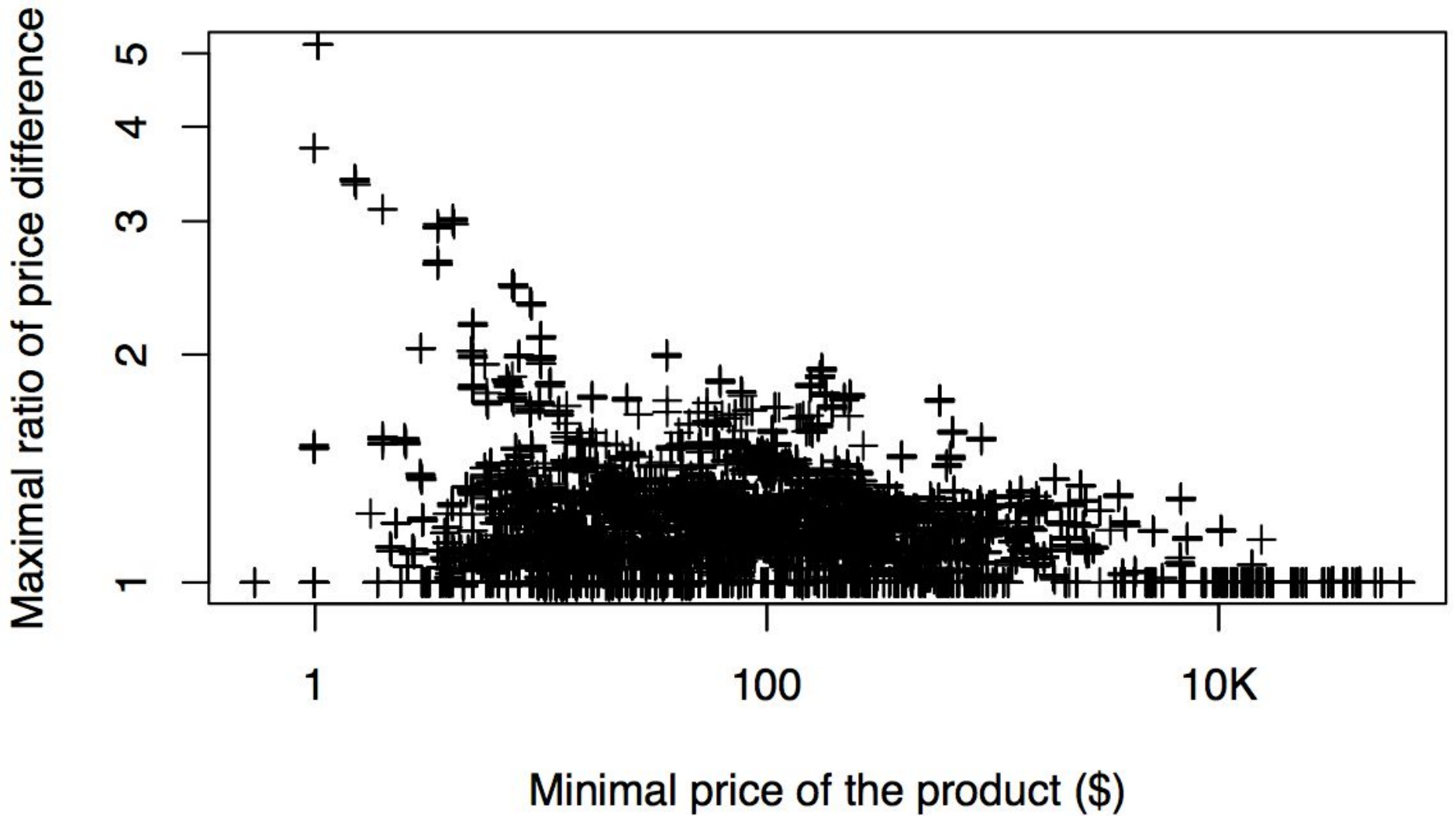
Which retailers?

ACM CoNEXT 2013

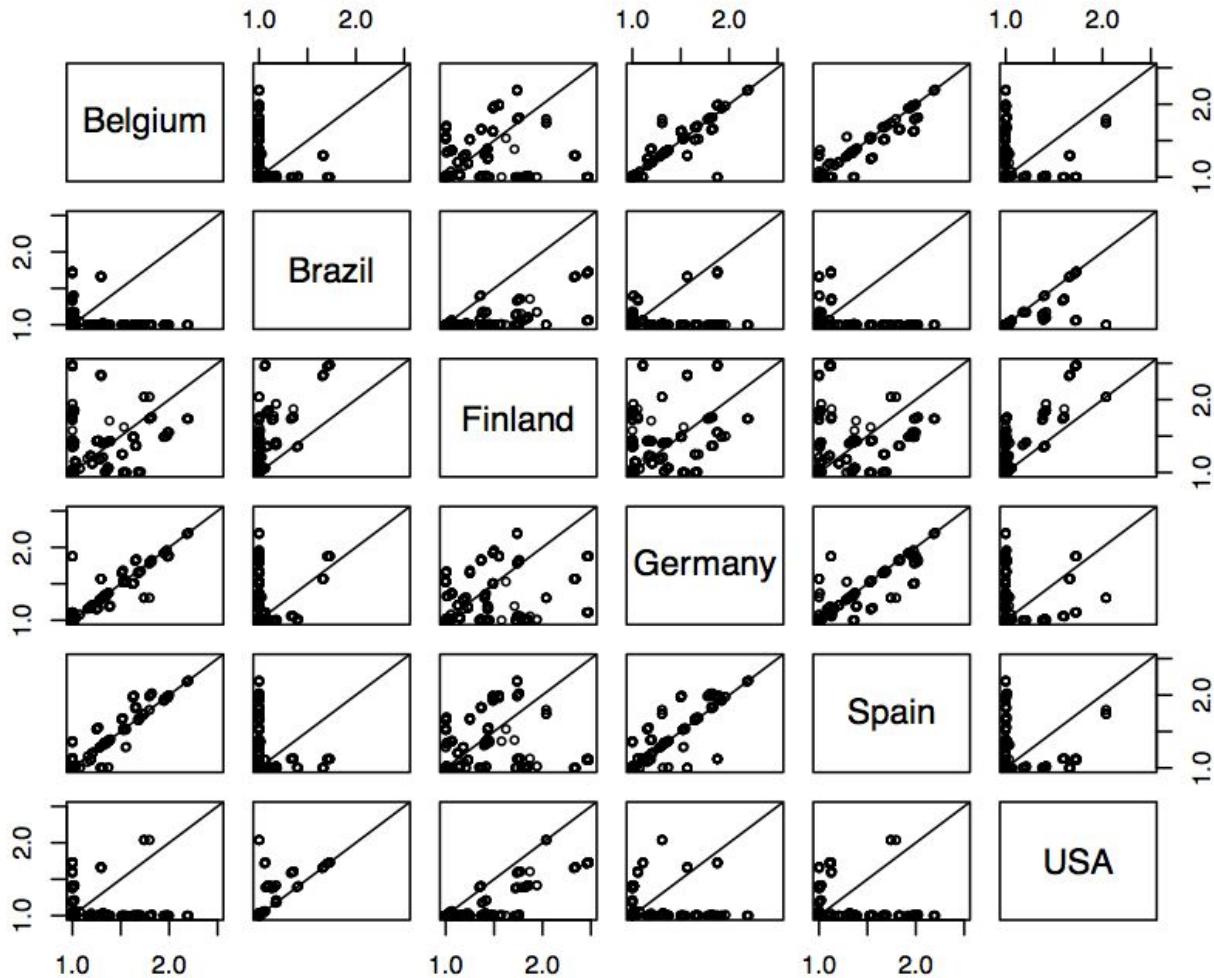
December 9-12, 2013
Santa Barbara, California



Which products?



Which countries?



(b) `www.amazon.com`

Detection of personalized PD



Variant	Converted Value	Original Text
You	€ 1378.12	€ 1,378.12
Windows 7, Chrome, Spain	€ 1378.12	€ 1,378.12
Mac OS, Safari, Spain	€ 1378.12	€ 1,378.12
Linux, Firefox, Spain	€ 1378.12	€ 1,378.12
United States, Tennessee	€ 1147.83	US\$ 1,299.00
United States, Washington	€ 1147.83	US\$ 1,299.00
Canada, British Columbia	€ 1274.79	C\$ 1,798.99
Canada, Ontario	€ 1274.79	C\$ 1,798.99
Canada, Ontario	€ 1274.79	C\$ 1,798.99
Israel, Beer-Sheva	€ 1411.83	US\$ 1,597.77
Sweden, Scandinavia	€ 1469.22	US\$ 1,662.72
Japan, Tokyo	€ 1205.22	US\$ 1,363.95
Japan, Hiroshima	€ 1205.22	US\$ 1,363.95
Czech Republic, Praha	€ 1377.40	US\$ 1,558.80
Korea, Seoul	€ 1492.18	US\$ 1,688.70
New Zealand, Dunedin	€ 1608.01*	NZ\$ 2,398.99

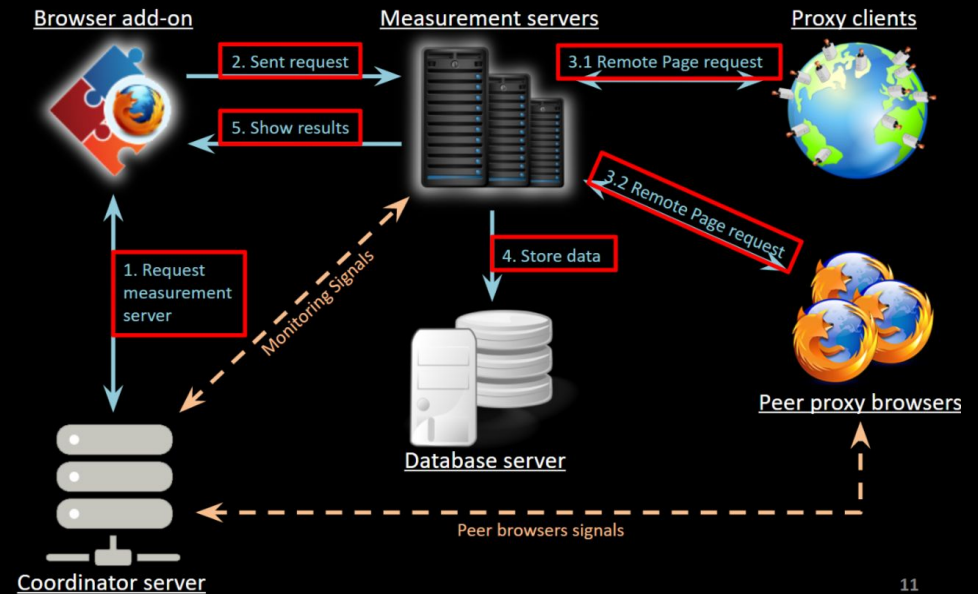
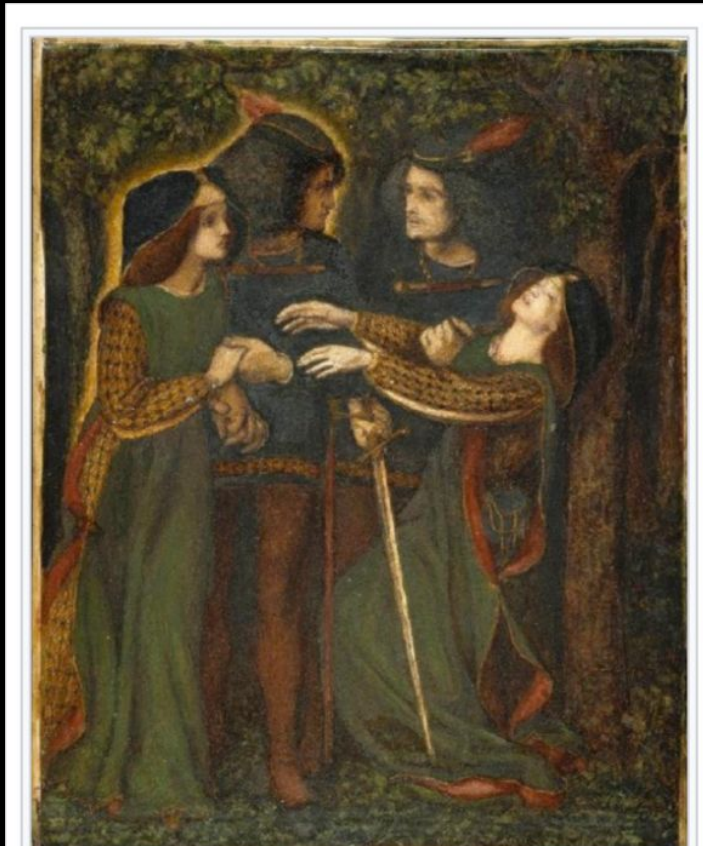
Results from local users

Source ID	Converted Value	Original Text
Local User 0	€ 1378.12	€ 1,378.12
Local User 1	€ 1378.12	€ 1.378,12

Do P2P checks



Solving the profile pollution problem



11

Doppelgänger

From Wikipedia, the free encyclopedia

For other uses, see [Doppelgänger \(disambiguation\)](#).

A **doppelgänger** (/ˈdɒpəlˌɡɛŋər/ or /-ɡæŋər/; German: [ˈdɔpəlˌɡɛŋɐ] listen (help·in·info)), literally "double-goer") is a **look-alike** or double of a living person, sometimes portrayed as a ghostly or **paranormal** phenomenon and usually seen as a **harbinger** of bad luck. Other traditions and stories equate a doppelgänger with an **evil twin**. In modern times, the term **twin stranger** is occasionally used.^{[1][2]}

Enough with PD ... lets get to tracking & advertising



What we search – The sites we visit – Who we befriend – What we buy ... everything is tracked

How can you tell if an ad is targeted?



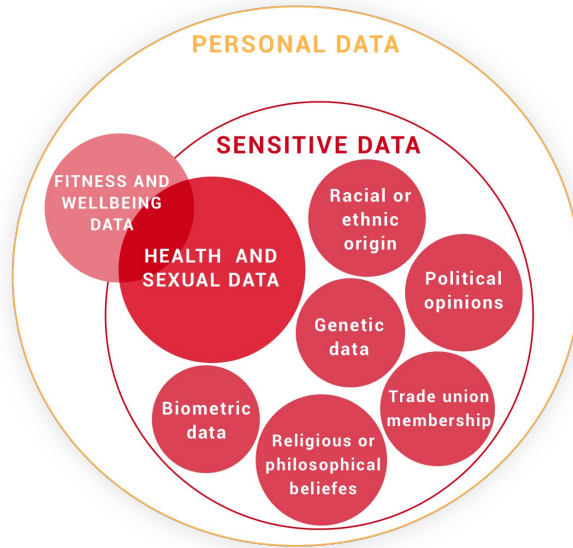
Tennis For Life

Sponsored • ⚙️

Limited Edition - Not found in stores
Order here => <https://tinyurl.com/yyqg7ntv>
Worldwide Shipping



This goes beyond curiosity ...



THE VERGE

TECH ▾ REVIEWS ▾ SCIENCE ▾ CREATORS ▾ ENTERTAINMENT ▾ VIDEO ▾ MORE ▾

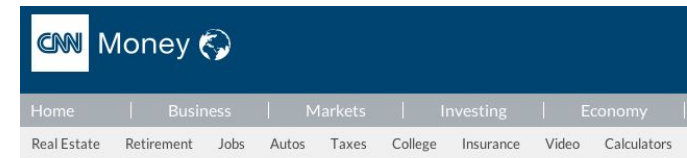
GOOGLE POLICY TECH

Google will pay \$170 million for YouTube's child privacy violations

It's the largest COPPA fine in history

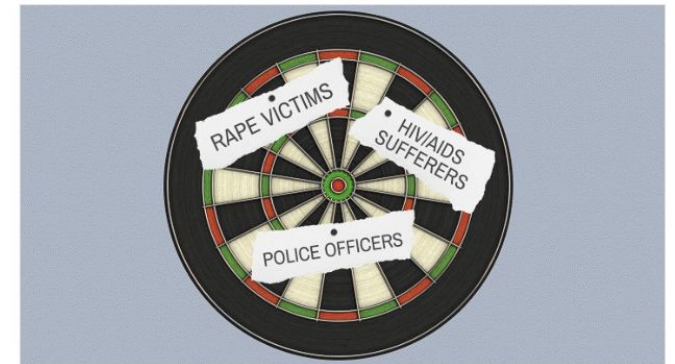
By Makena Kelly | @kellymakena | Sep 4, 2019, 9:41am EDT

f t SHARE



Data brokers selling lists of rape victims, AIDS patients

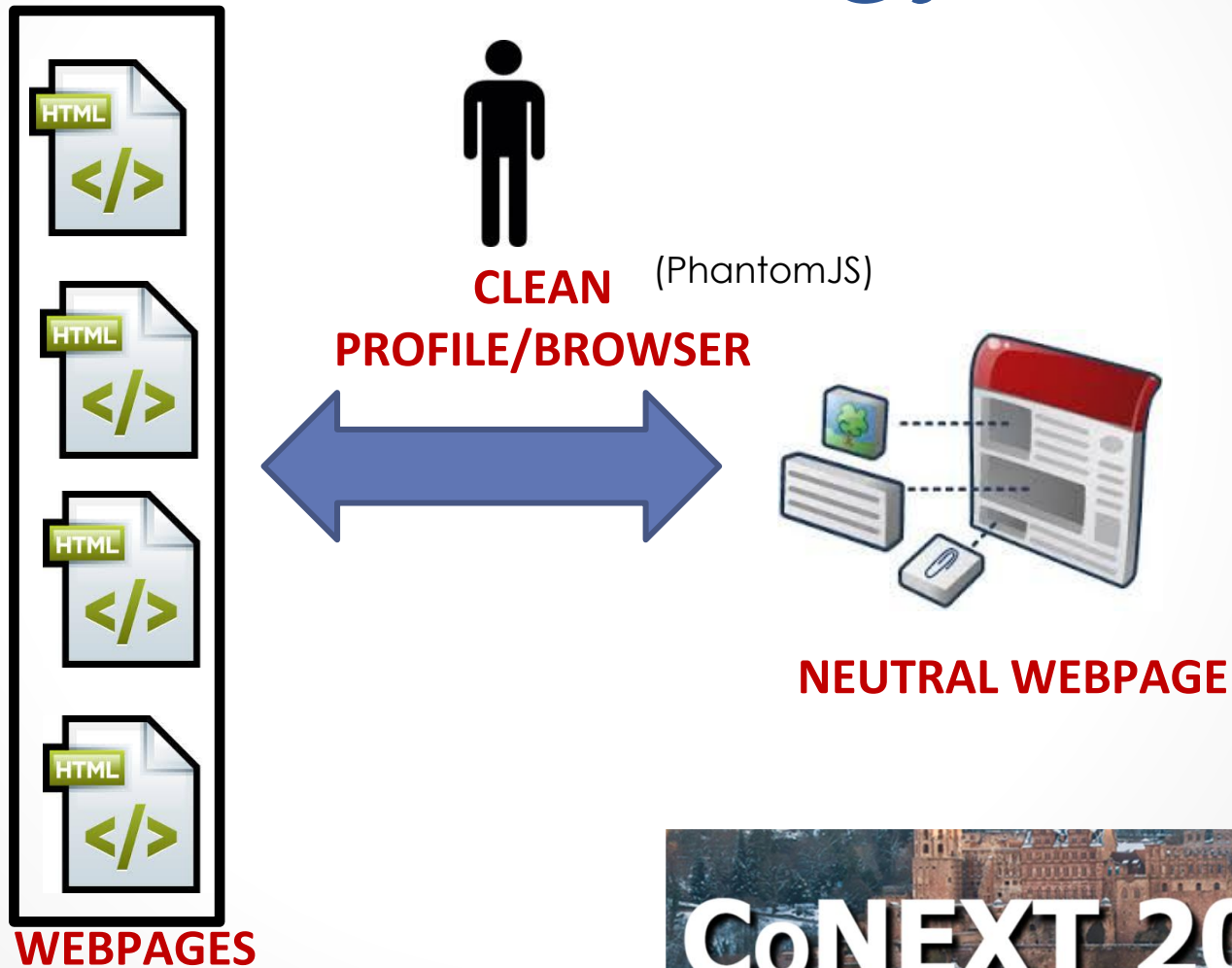
By Melanie Hicken @melhicken December 19, 2013: 12:38 PM ET



First approach

Detection via content-based analysis

Main idea of our methodology

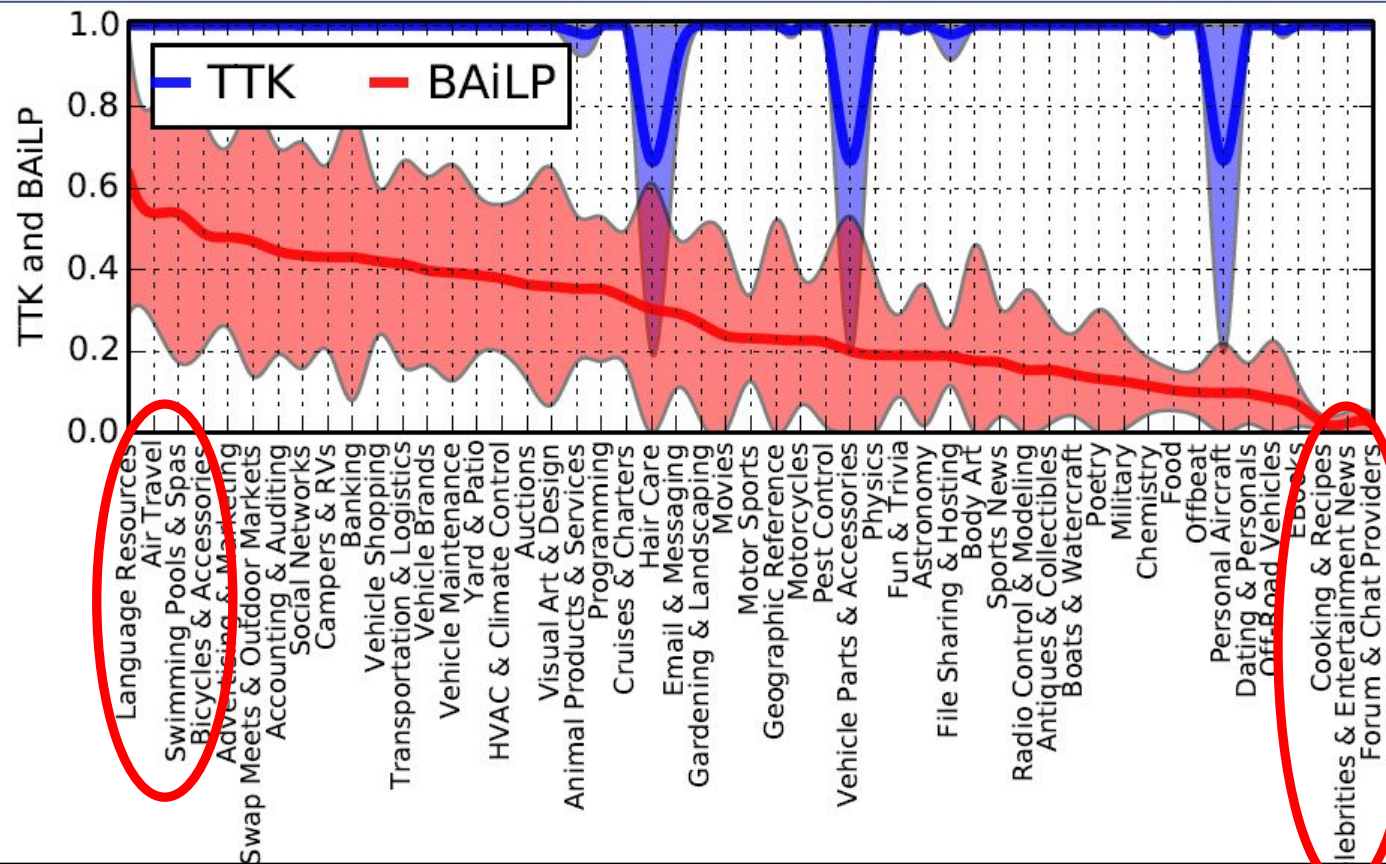


How frequent is OBA?

Are some personas more targeted than others?

In summary,

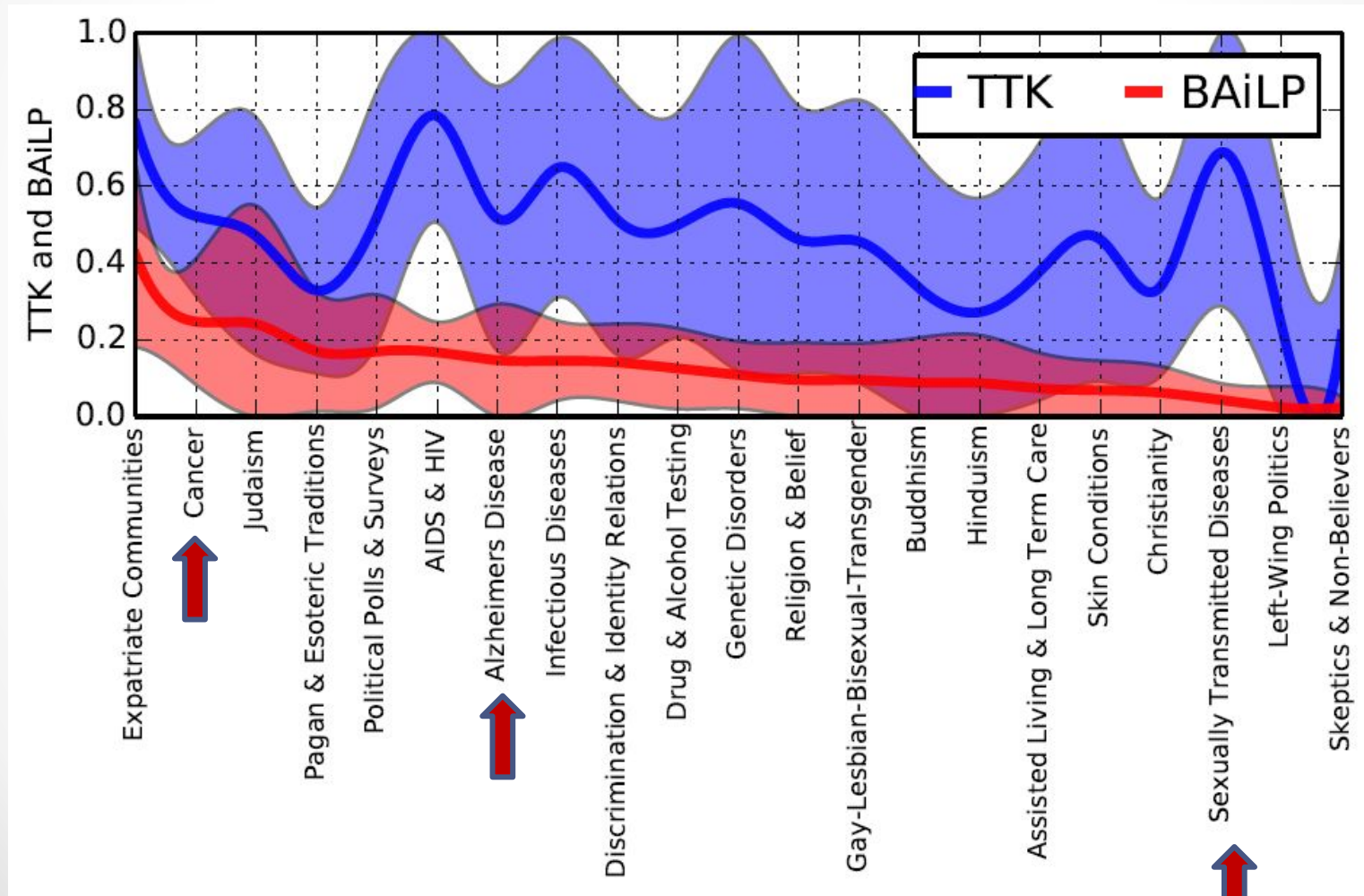
- TTK measures if OBA is happening
- BAiLP captures what percentage is due to OBA.



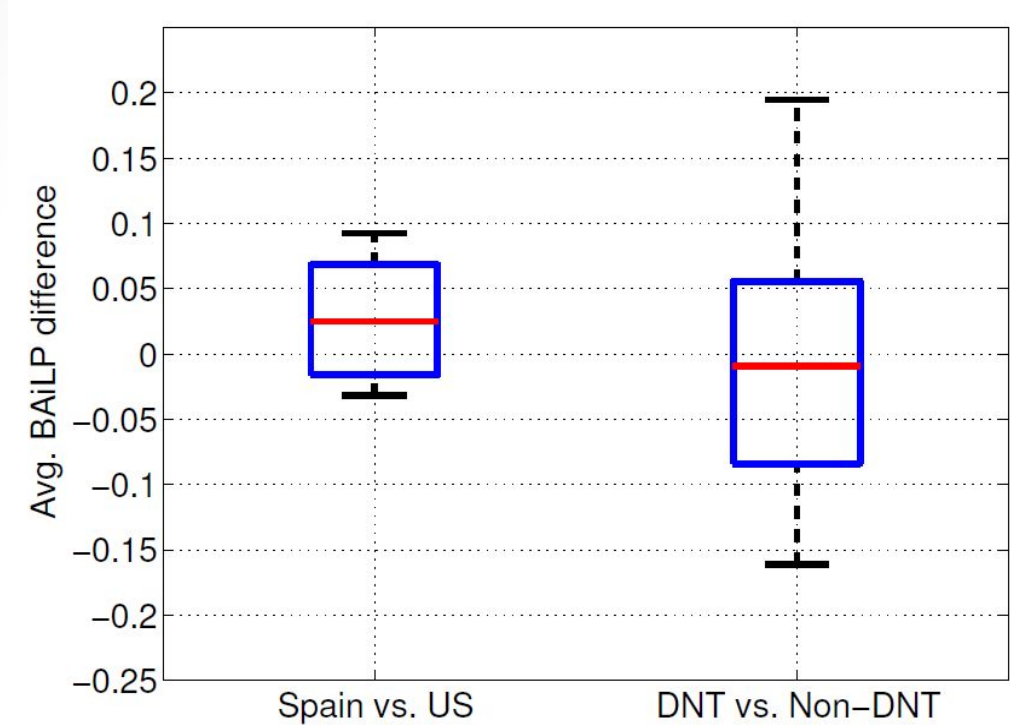
• **CORRELATION** → received OBA (BAiLP) and its value for the advertisers • 22

Is OBA applied to sensitive topics?

- Same methodology → 21 sensitive *personas*



Geographical and Do-Not-Track



- **Same methodology:**
 - Geographical: US and ES
 - Do-Not-Track: ON | OFF

Limitations of content-based analysis detection

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



2016 ELECTION | NOV. 22, 2016

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

By Halle Kiefer

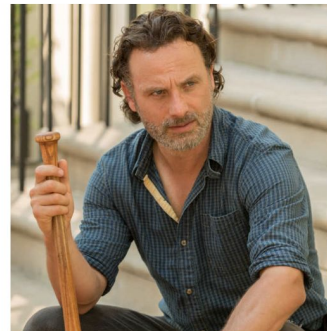


Photo: Gene Page/AMC

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:

Second approach

Count-based detection & crowdsourcing

Targeted ads follow you around

- Detection via simple counting
- No need for content analysis
- No need to inject traffic
- Real-time
- No prob with indirect targeting



Detects & annotates all rendered ads



eyeWnder Ads Analyzer

Tools & Options

Browsing History Tools

eyeWnder Options

Detected Ads are annotated with the following images. Click when you see one to get more info about the Ad.

Note that eyeWnder Ad detection is not working when you are also using an Ad Blocking extension.

0% 100%

Training Status: 100%

For more information click [here](#) - Ver. 0.1.3

WIRED

CULTURE

DESIGN

GEAR

SCIENCE

SECURITY



KALOAD Z40 Smart Watch



Banggood.com

Lowest Price
\$ 29.99

WIRED

CULTURE

DESIGN

GEAR

SCIENCE

SECURITY



¿Debería comprar acciones ahora?

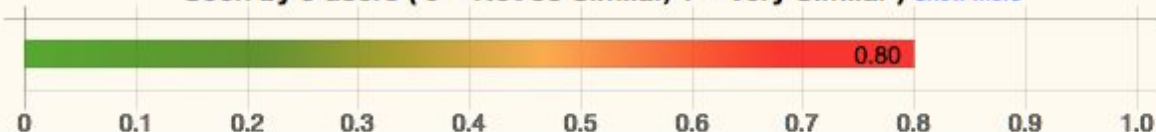
Si tiene una cartera de valores igual o superior a 350.000 €, y quiere saber más sobre la dirección del mercado de valores, lea nuestra guía "Perspectivas de los mercados".

FISHER INVESTMENTS ESPAÑA®

¡Descubra más aquí!

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](#)**Advertisement Information Table**

This ad takes you to:	https://subscribe.wired.com/subscribe/wired/103100?source=WIR_Footer_IntlTargeting_Apr16
This ad is about:	<ul style="list-style-type: none"> People & Society > Men's Interests Books & Literature > Magazines News > Technology News

[Analyze Advertisement](#)[Get Intermediaries](#)

eyeWnder ©

A simple algorithm

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

Require:

Counters:

$\#Users_{\alpha}$ \triangleright Number of other users that observe ad α

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

Thresholds:

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

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

1: **if** $\#Users_{\alpha} \leq Users_{th}$ **AND** $\#Domains_{u,\alpha} \geq Domains_{th,u}$
then

2: Targeted ad

3: **else**

4: Non-targeted ad

Works pretty fine

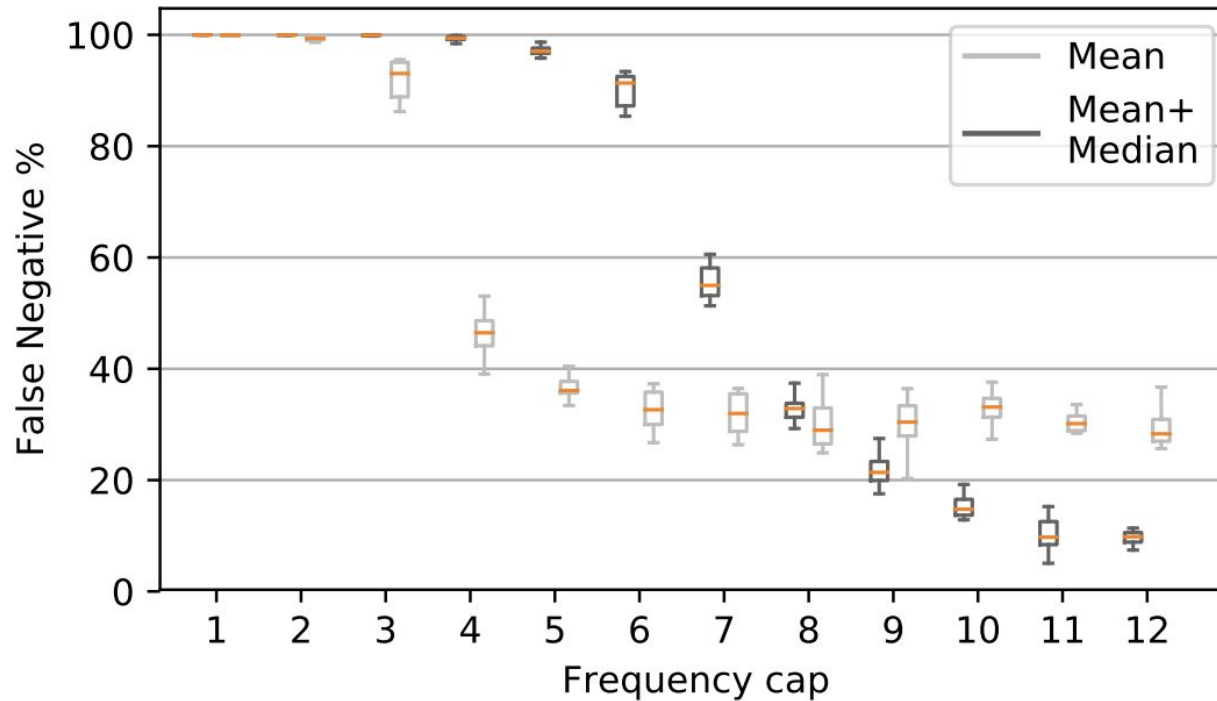
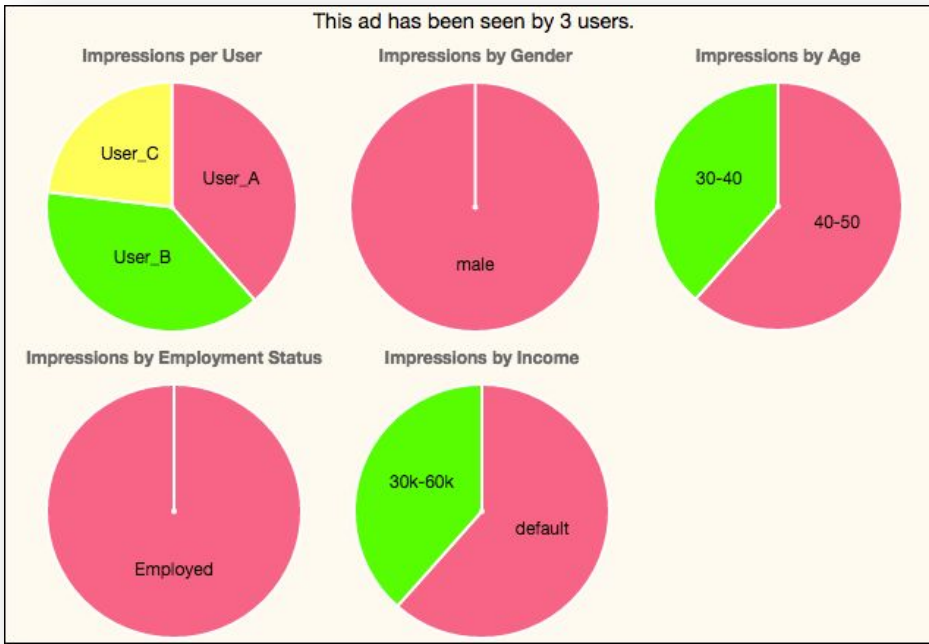
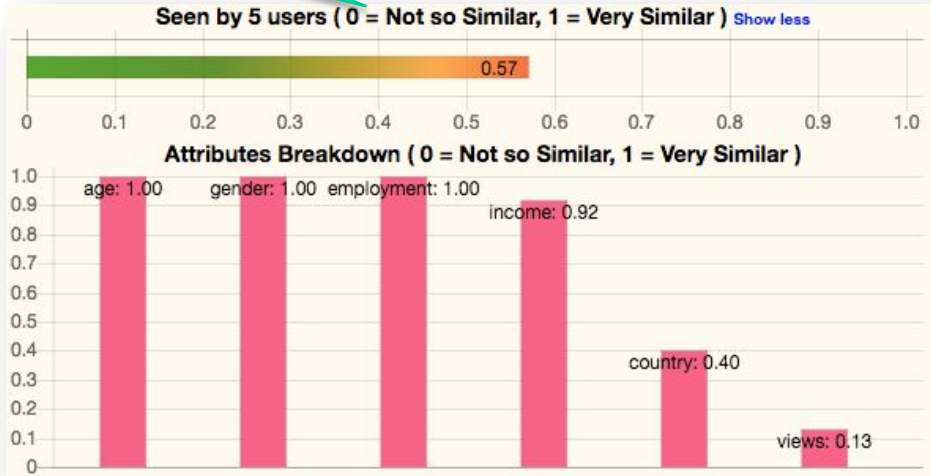


Figure 3: False Negatives % Vs. Frequency Cap using two different thresholds (Mean, Mean+Median) for both variables ($\#Users_{\alpha}$, $\#Domains_{u,a}$)

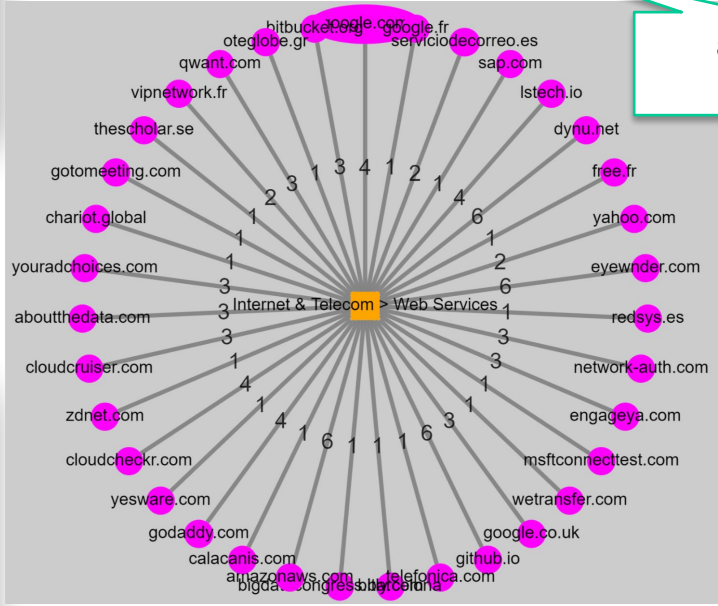


Similarities with other users

Crowd-sourced Users statistics



adwords cloud



Per Country Ad Impressions

Launch of Data Transparency Lab



Telefonica



mozilla

DATA TRANSPARENCY LAB

A community of technologists, researchers, policymakers and industry representatives working to advance online personal data transparency through scientific research and design.

Kick-off Workshop : Nov'14, Bcn

Participants included:

Northeastern University, MIT Human Dynamics Lab, Microsoft, Telefonica Innovation, Max Planck Institute for Software Systems, Mozilla, and more.



Downloads:

[DTL2014 Workshop Summary Report](#)

[DTL Vision](#)

[DTL Organization](#)

Building a community

Data Transparency Lab Call For Proposals April 2015

The Data Transparency Lab is a collaborative effort between universities, businesses and institutions to support research in tools, data, and methodologies for shedding light on the use of personal data by online services, and to empower users to be in control of their personal data online. In order to support research in these areas, DTL will award research grants to academic institutions worldwide. Such grants come in the form of a lump sum of up to 50K euro that is awarded to successful applicants for pursuing DTL related research in any of the following topics:

Topics

Tools, Platforms, Measurements, and Methodologies for:

Reverse-Engineering Personal Data Usage in Online Services (e.g., advertising, recommender services, pricing and availability of goods & information):

- Behavioral targeting
- Context / Location-based targeting
- Social graph-based targeting
- Involuntary (or implicit) customization / targeting

Detecting Personal Data Gathering by Online Services:

- Techniques for finger-printing and tracking users
- In-app tracking and targeting
- Information leakage from applications and platforms
- Cross-platform/domain information/profile trading, aggregation, fusion

Privacy-preserving Personal Data Analytics/Management:



18 grants & lots of collaborations



DATA TRANSPARENCY LAB - EXAMPLE 1: FACEBOOK DATA VALUATION TOOL

Permits users to estimate how much money Facebook is making on them

INFORMS INDIVIDUAL USERS



A plugin for your browser that combines your online activity with Facebook's Public APIs to estimate your advertising value

DEMO VIDEO: <https://youtu.be/QPfc-gXGdjl>

REVEALS SOME MORE GENERAL TRENDS



Un español vale la mitad que un americano en Facebook

Una investigación realizada por dos hermanos españoles mide cómo la red social pone precio a los anuncios de los perfiles

Demonstrates how factors like country, status, studies, etc. impact on a user's advertising value

LIVE DEMO

https://acrumin.cartodb.com/viz/75d6d052-0648-11e6-8923-0e3ff518bd15/public_map

DATA TRANSPARENCY LAB - EXAMPLE 2: PRIVACY

CENSUS

Privacy Census

A CENSUS THAT IDENTIFIES WEBSITES THAT TRACK USERS

Sites with canvas fingerprinting scripts

In a crawl conducted during January 2016, these websites were found to run scripts on their homepages that used the Canvas API to fingerprint users.

Show 25 entries Search:

Showing 1 to 25 of 15,089 entries

Alexa Rank	Site URL	Fingerprinting Domain
11	http://taobao.com	alicdn.com
29	http://tmall.com	alicdn.com
97	http://dropbox.com	dropboxstatic.com
115	http://bbc.com	doubleverify.com
143	http://cnzz.com	tbcn.cn
153	http://detail.tmall.com	alicdn.com
178	http://avito.ru	avito.st
219	http://washingtonpost.com	doubleverify.com

Privacy Auditor

You are being tracked.

[dropbox.com](#) has been spotted during the last month using the following fingerprinting techniques:

- ! Canvas Fingerprinting

Check my browser history vs. fingerprinting

Share your results:

Twitter Facebook

What is fingerprinting?

About this tool

TO SHED LIGHT ABOUT HOW TRACKING IS USED AND BY WHOM

MIT Technology Review

Computing

Largest Study of Online Tracking Proves Google Really Is Watching Us All

Google's Web trackers are present on the majority of the Web's top million sites.

by Tom Simonite May 18, 2016



Audio fingerprinting being used to track web users, study finds

Posted May 19, 2016 by [Natasha Lomas](#) (@riptar)

Data transparency must combine cutting-edge research with community involvement

Arvind Narayanan, Assistant Professor of computer science at Princeton.

26 May 2016

[Key to the success of data transparency...](#)

Steve Englehardt and I recently made available our draft paper [Online tracking: A 1-million-site measurement and analysis](#), funded in part by a DTL grant. It is part of the [Web Transparency & Accountability Project](#) at Princeton, and it's the most detailed look at online tracking conducted so far. Among our findings was the fact that the the Audio, Battery, and WebRTC APIs in HTML5 are all being abused by third-party scripts for fingerprinting. There's been some press coverage [here](#) and [here](#).

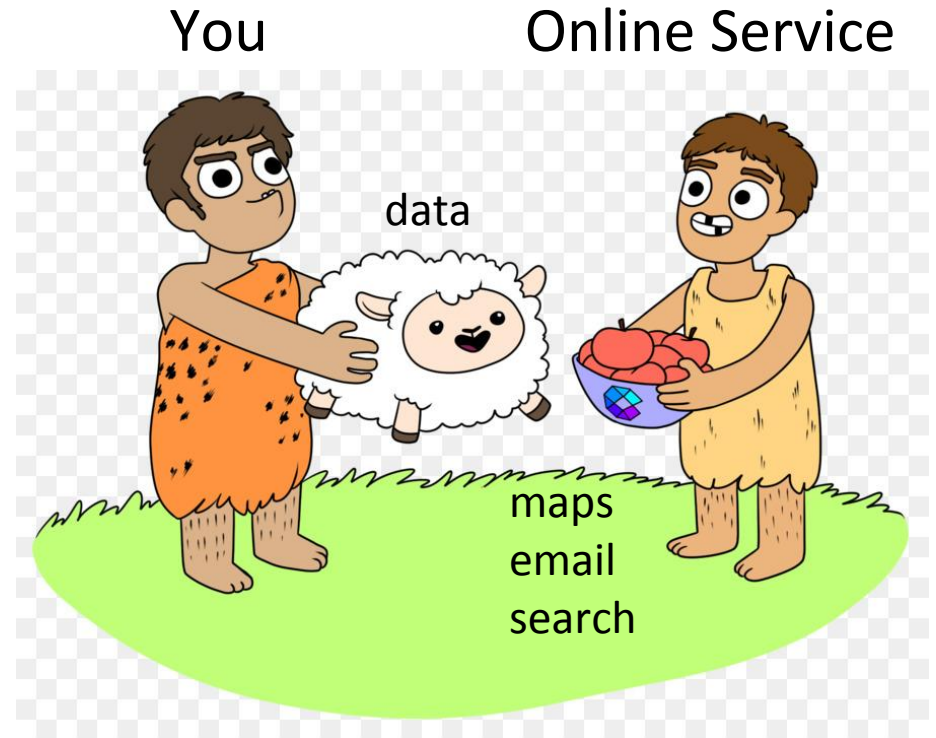
Big Idea #2 - NOT obvious even in retrospect

Most problems of the web are due to its broken economics model



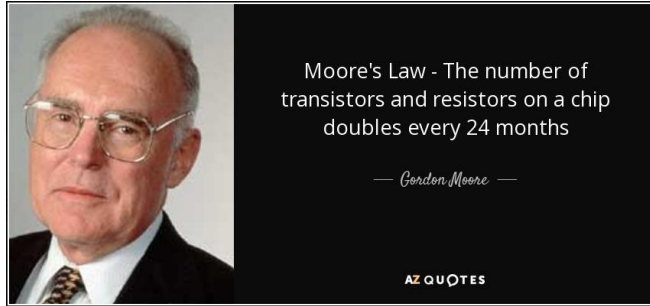
You and online services

- Free data
- For free service
 - payment “in kind”
 - no cash

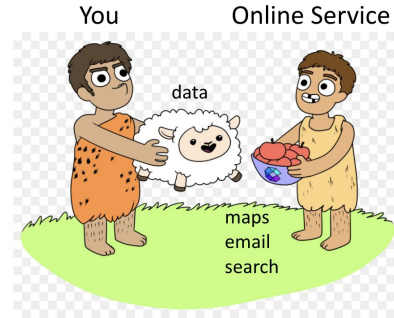


BAD for privacy!

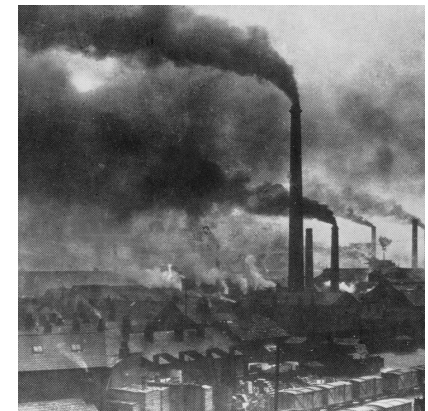
(tracking is cheap)



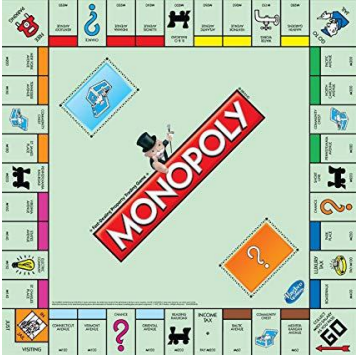
(no payments to users)



(collect ... everything)

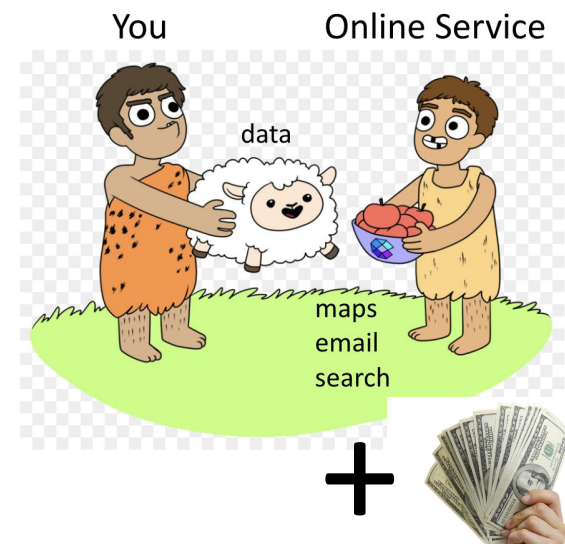


BAD for sustainability!

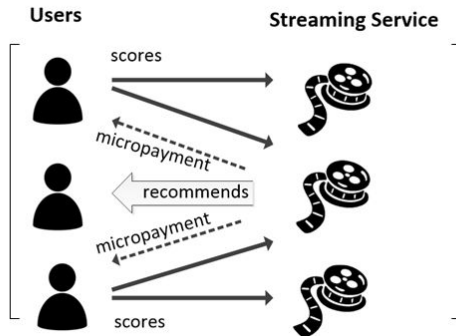


A Human-Centric Data Economy

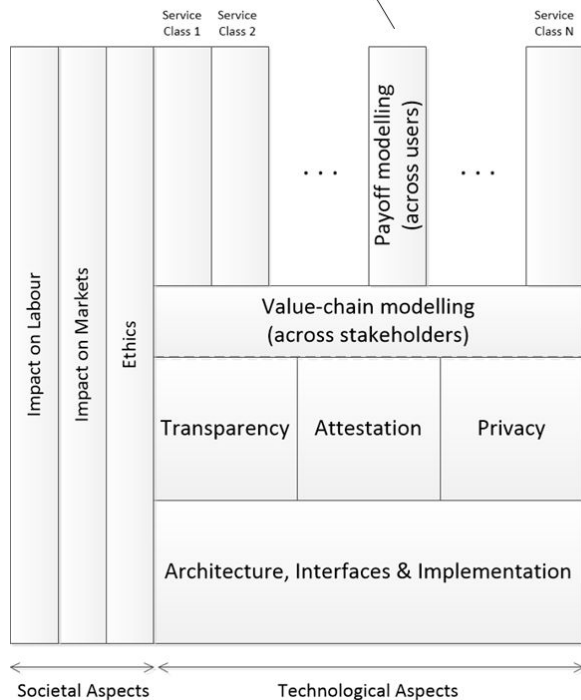
- explicit monetary compensation for data based on their value for online services
 - e-commerce
 - media streaming
 - location services
 - ...



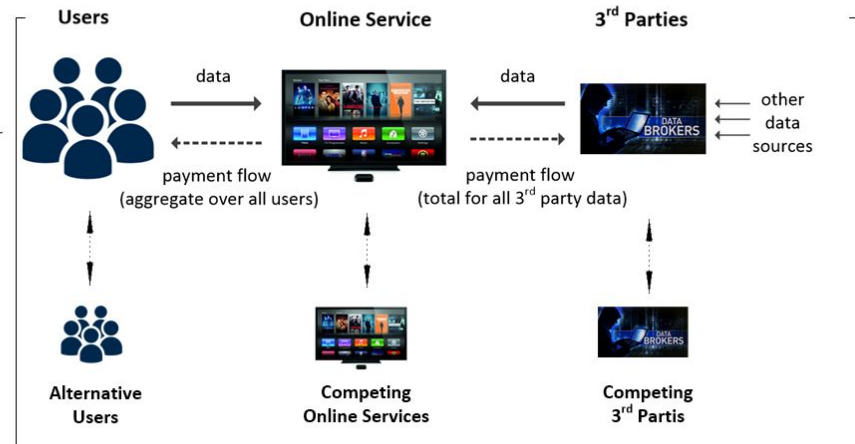
Tons of great questions to ask



How to split the payoff corresponding to a single transaction among the users that provided data for this transaction



How to split total value between the different stake holder types: Users, Online Services, 3rd Party Data Providers



Modelling and Algorithms

Accounting and Implementation

HCDE vs. Data Marketplaces 1/4

Hundreds of DMs

- Aggregate
- Personal Data

Data pricing

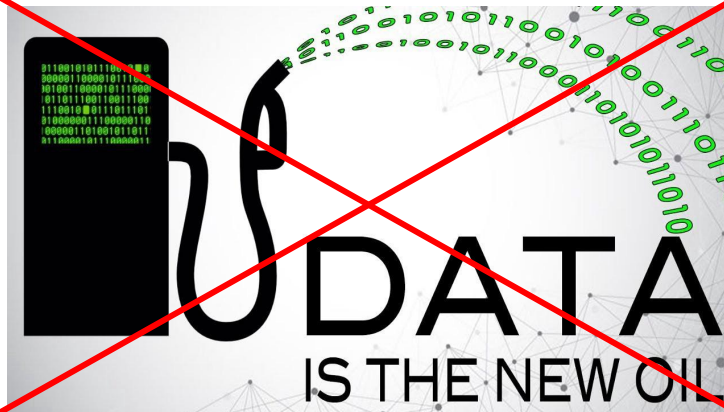
- Ad hoc
- Auction-based

<p>General</p> <p>Airbloc personal data</p> <p>Decentralized</p>	<p>General</p> <p>BIGToken personal data</p> <p>Decentralized</p>	<p>General</p> <p>Bron.Tech personal data</p> <p>Decentralized</p>
<p>General</p> <p>Canada Open Data Exchange open data</p>	<p>General</p> <p>DACONOMY no specification</p> <p>Decentralized</p>	<p>General</p> <p>DAX personal data</p> <p>Decentralized</p>
<p>General</p> <p>Data Blockchain no specification</p> <p>Decentralized</p>	<p>General</p> <p>Data Markets healthcare data, personal data, alternative data</p>	<p>General</p> <p>Datacoup Personal Data Marketplace personal data</p>
<p>General</p> <p>Dataeum location data</p> <p>Decentralized</p>	<p>General</p> <p>Datareum personal data</p> <p>Decentralized</p>	<p>General</p> <p>Datatrade personal data</p>
		<p>General</p>

Source: <https://about.datarade.ai>

HCDE vs. Data Marketplaces 3/4

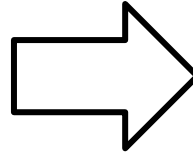
- Data is **not** really a commodity
- Two different liters of oil are almost identical
- But what about
 - Browsing behavior famous investor vs. average person?
 - Mobility data from a taxi driver vs. a weekend driver?
 - Shopping cart of a teenager vs. middle-ager?



HCDE vs. Data Marketplaces 4/4

Digital goods / Information

- Don't decay
- Cost to copy = 0
- Data provenance = hard



Without Clear Context

- Hard to price
- Hard to auction



X euros

- Infinite days
- Infinite Km
- Infinite drivers

- * Consumption 0 lt per 100Km
- * Service every infinite years

To probe further

Valuating User Data in a Human-Centric Data Economy

Marius Paraschiv
IMDEA Networks
Leganés – Madrid

marius.paraschiv@imdea.org

Nikolaos Laouraris
IMDEA Networks
Leganés – Madrid

nikolaos.laouraris@imdea.org

Feature Article: Data Economy

Why Online Services Should Pay You for Your Data? The Arguments for a Human-Centric Data Economy

7 Nikolaos Laouraris
8 IMDEA Networks Institute

9 **Abstract—Data, and the economy around it, are said to be driving the fourth industrial**
10 **revolution. Interestingly, the people—whose data are what moves the new economy—**
11 **have a rather passive role in it as they are left outside the direct value flow that transforms**
12 **raw data into huge monetary benefits. This is a consequence of a de facto understanding**
13 **(or, one may say, misunderstanding) between people and companies that the former**
14 **receive unpaid access to online services in exchange for the unpaid access to their**
15 **personal data. This article argues in favor of an alternative human-centric data economy**
16 **in which people will be paid whenever their data will be used by revenue-generating**
17 **products and services. We discuss the benefits of such an economy, the main challenges**
18 **for realizing it, and its feasibility in the view of existing technologies and business**
19 **practices.**

20 ■ **IMAGINE A FUTURE** in which a recommendation hotel room at a booking website, or for a movie
21 for a product at an e-commerce website for a at a pay-per-view streaming service, would all
redistribute a part of the resulting payment
among the users whose previous shopping,
travel, or viewing patterns were mined in order

Digital Object Identifier 10.1109/MIC.2019.2953764

XXXX 2019

Published by the IEEE Computer Society

1089-7801 © 2019 IEEE

1

Abstract—The idea of paying people for their data is increasingly seen as a promising direction for resolving privacy debates, improving the quality of online data, and even offering an alternative to labour-based compensation in a future dominated by automation and self-operating machines. In this paper we demonstrate how a Human-Centric Data Economy would compensate the users of an online streaming service. We borrow the notion of the Shapley value from cooperative game theory to define what a fair compensation for each user should be for movie scores offered to the recommender system of the service. Since determining the Shapley value exactly is computationally inefficient in the general case, we derive faster alternatives using clustering, dimensionality reduction, and partial information. We apply our algorithms to a movie recommendation data set and demonstrate that different users may have a vastly different value for the service. We also analyse the reasons that some movie ratings may be more valuable than others and discuss the consequences for compensating users fairly.

1. INTRODUCTION

Data, and the economy around it, are said to be driving the fourth industrial revolution. Interestingly – the people – whose data is what moves the new economy, have a rather passive role in it, as they are left outside the direct value flow that transforms raw data into huge monetary benefits. This is a consequence of the de facto understanding (or one may say misunderstanding) between people and companies, that the former get unpaid access to online services in exchange for unpaid access to their personal data. This is increasingly being challenged by various voices who call for the establishment of a new, renegotiated, relationship between users and services. Indeed, a variety of pathologies can be traced back to the way the data economy has been working so far. Some are direct and obvious, such as privacy risks for individuals, and market failures and dangers for the economy from the rise of data monopolies and oligopolies. Others are less obvious, and further reaching into the future, such as mass unemployment due to data-driven automation.

It was estimated recently [1] that, if automation due to artificial intelligence reaches maturity and fair remuneration algorithms are set in place, a family of four could earn up to \$20,000 per year from their data. The idea of micropayments, or providing small contributions to users in exchange for their presence on a platform or for accessing a service, is of course much older. In the pre-World Wide Web era, France developed a videotex online service called Minitel, that included micropayments as part of its design, but Jaron Lanier brought it to public attention in 2013, in his book “Who owns the future?” [2]. In it, he argues that we have only underwent half of the Data Revolution, the part that

compensates users with *implicit* benefits, but not the part that also compensates them with *explicit* monetary benefits.

There have been a series of proposed approaches for how this compensation might materialise. The simplest, at least in theory, would be to assign a context-free value to data, a kind of dollar-per-bit measure. This has been proven to be very hard [3], [4], [5], [12]. Indeed, since the value of data is strongly connected to its intended use, it becomes very difficult to argue about how to assign an a priori average value. For traditional currencies, we are able to have a context-free appreciation of their value for the simple reason that we have been using these currencies long enough to be able to do so. Although we clearly understand nowadays that one’s browsing and mobility patterns, social network, or past purchases all have value, we are far from being able to appreciate how much this value is in terms of dollars or euros. The latter is further complicated by our inability to tell in advance, by how many parties, and how many times, a piece of data may be utilized. As an analogy, selling an individual’s data, or rather renting it temporarily, is as difficult and risky as renting an infinitely fast vehicle, with no gas and maintenance costs, and without any prior restrictions with regard to mileage or the person driving it.

A second proposed method has been to compensate users for their privacy damage [13], [14]. Processing massive amounts of data can lead to privacy infringements, such as the leakage of habitual user behavior, their location or other personal identifiable information (PII). Users are thus seen as victims who must be compensated for their damage.

Our approach is different, we consider users as active partners in the data value chain. Such a chain requires a business model, smart predictive algorithms for extracting useful information from raw data and online marketing for attracting and retaining users, among many others. The fundamental component of the value chain, however, is the user, and it is ultimately a matter of common sense that they should be rewarded in a fair manner, which may or may not exceed the perceived privacy-related damages.

In a Human-Centric Data Economy, when a transaction, or set of transactions, is converted, a proportion of the obtained revenue will be returned to the users. Defining the right amount to be returned to the users is difficult, as it depends on many market characteristics of a multilateral value chain, such as competition and user loyalty [6]. In this paper, we assume that the total amount of revenue to be redistributed to users is given, e.g., 5% or 10%, or any other number produced by the competition between services

Almost done



people don't care about privacy
(some say)

some other things people didn't care about

smoke



flight security



kids playing with melted glass



kids playing with melted iron



kids playing with power tools



still available at ebay

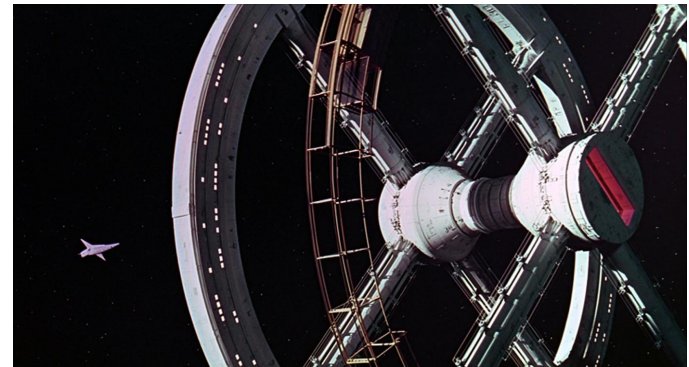
kerosene train



haven't located one yet



societies evolve



Won't be long before we look back and
shake our head



Thank you!

