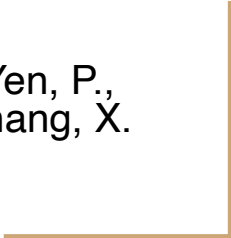# Network Structure of the Digital Advertising Marketplace

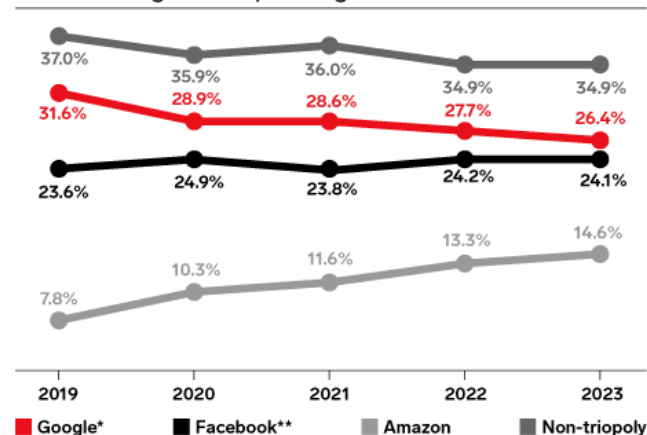Subramanian, V., Chun-Yen, P., Stukalin, S., Sharp, M., Zhang, X.

The Problem - what is it and why is it important

- The Problem
  - Perceived market power in digital advertising

- Urgency
  - Bipartisan action already underway based on the perception of market power
    - *Competition and Transparency in Digital Advertising Act* 2022
  - Breakdown vertical integration

- Antitrust legislation

US Triopoly Digital Ad Revenue Share, by Company, 2019-2023

% of total digital ad spending

	2019	2020	2021	2022	2023
Non-triopoly	37.0%	35.9%	36.0%	34.9%	34.9%
Google*	31.6%	28.9%	28.6%	27.7%	26.4%
Facebook**	23.6%	24.9%	23.8%	24.2%	24.1%
Amazon	7.8%	10.3%	11.6%	13.3%	14.6%

■ Google* ■ Facebook** ■ Amazon ■ Non-triopoly

*Note: includes advertising that appears on desktop and laptop computers as well as mobile phones, tablets and other internet-connected devices, and includes all the various formats of advertising on those platforms; net ad revenues after companies pay traffic acquisition costs (TAC) to partner sites; *includes YouTube advertising revenues; **includes Instagram advertising revenues*
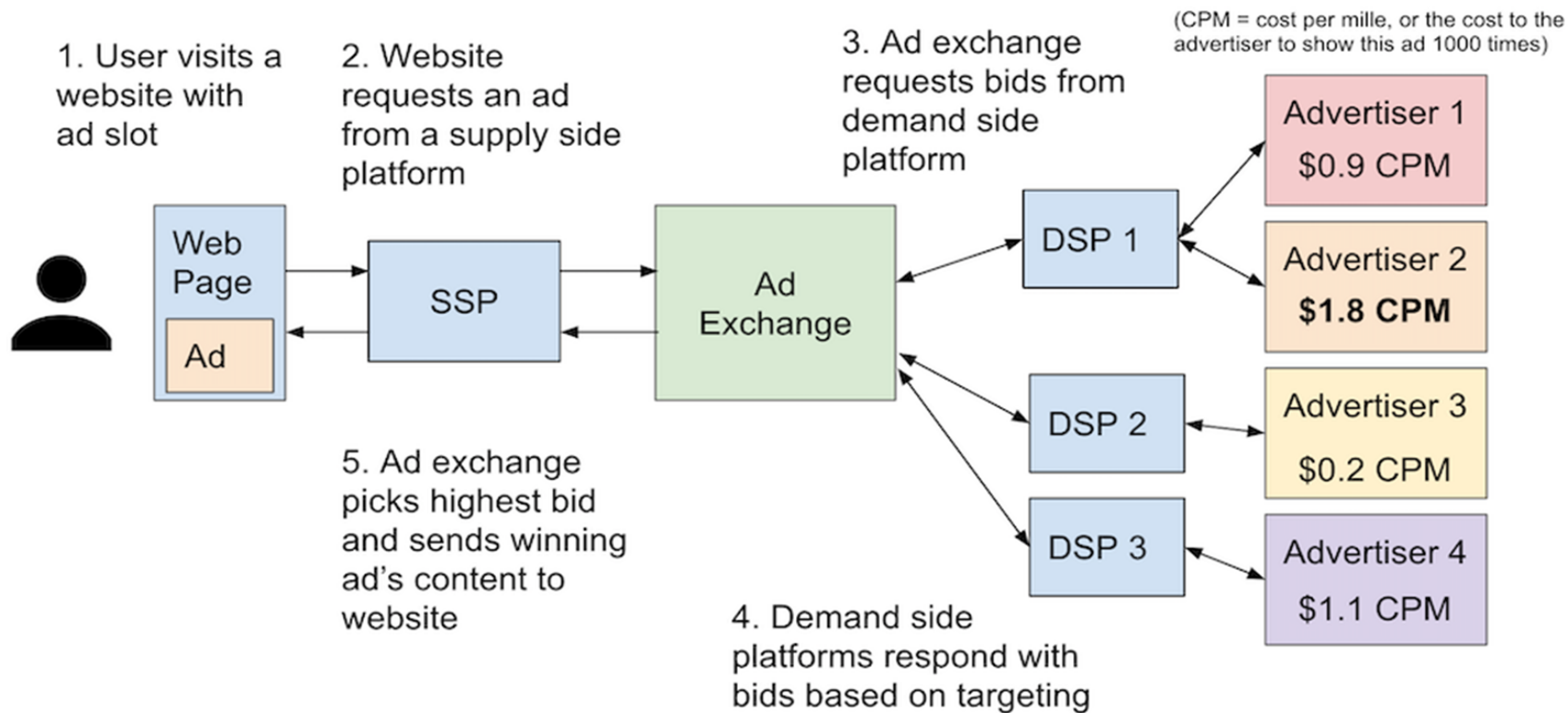*Source: eMarketer, Oct 2021*

270192   eMarketer | InsiderIntelligence.com

Image taken from https://www.insiderintelligence.com/content/google-facebook-amazon-account-over-70-of-us-digital-ad-spending
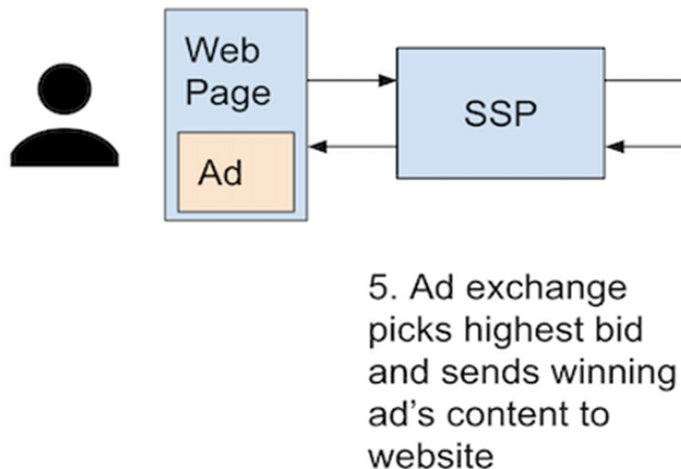
Domain Knowledge

1. User visits a website with ad slot

2. Website requests an ad from a supply side platform

3. Ad exchange requests bids from demand side platform

(CPM = cost per mille, or the cost to the advertiser to show this ad 1000 times)

Web Page

Ad

SSP

Ad Exchange

DSP 1

Advertiser 1
$0.9 CPM

Advertiser 2
$1.8 CPM

DSP 2

Advertiser 3
$0.2 CPM

DSP 3

Advertiser 4
$1.1 CPM

5. Ad exchange picks highest bid and sends winning ad's content to website

4. Demand side platforms respond with bids based on targeting

Our data

Data on the Supply side (ads.txt)
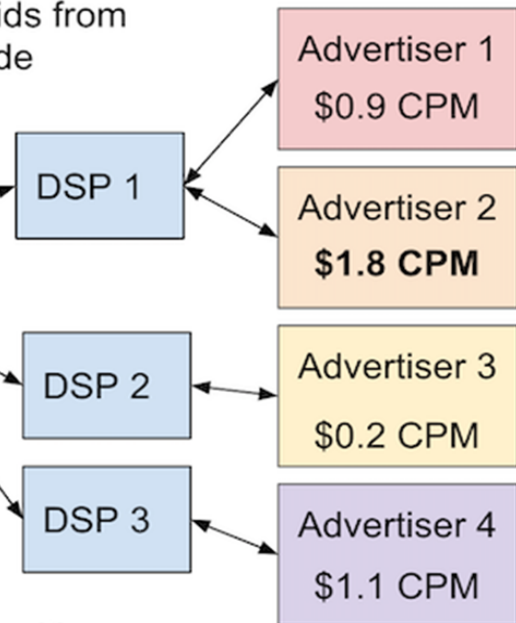
1. User visits a website with ad slot

2. Website requests an ad from a supply side platform

3. Ad exchange requests bids from demand side platform

(CPM = cost per mille, or the cost to the advertiser to show this ad 1000 times)

Web Page

Ad

SSP

Ad Exchange

DSP 1

DSP 2

DSP 3

Advertiser 1
$0.9 CPM

Advertiser 2
$1.8 CPM

Advertiser 3
$0.2 CPM

Advertiser 4
$1.1 CPM

5. Ad exchange picks highest bid and sends winning ad's content to website

4. Demand side platforms respond with bids based on targeting

Example dataset (https://nytimes.com/ads.txt)

```
amazon-adsystem.com, 3030, DIRECT
appnexus.com, 3661, DIRECT
google.com. pub-4177862836555934, DIRECT
google.com, pub-9542126426993714, DIRECT
indexexchange.com, 184733, DIRECT
liveintent.com, 130, DIRECT
openx.com, 537145107, DIRECT
openx.com, 539936340, DIRECT
openx.com, 539052954, DIRECT
openx.com, 544071378, DIRECT, 6a698e2ec38604c6
rubiconproject.com, 12330, DIRECT
rubiconproject.com, 17470, DIRECT
triplelift.com, 746, DIRECT
pubmatic.com, 158573, DIRECT, 5d62403b186f2ace
pubmatic.com, 158945, DIRECT, 5d62403b186f2ace
media.net, 8CU2553YN, DIRECT
yahoo.com, 55861, DIRECT, e1a5b5b6e3255540
yahoo.com, 55792, DIRECT, e1a5b5b6e3255540
google.com, pub-1793726897772453, DIRECT, f08c47fec0942fa0
aps.amazon.com, 3030, DIRECT
indexexchange.com, 196165, DIRECT, 50b1c356f2c5c8fc
adswizz.com, nytimes, DIRECT
triplelift.com, 746-EB, DIRECT, 6c33edb13117fd86
liveintent.com, 74445, DIRECT
```

Each line contains up to 4 fields (columns):

- **Ad network/SSP** (domain)
  - Informs seller that ad network is authorized network with publisher
- **Publisher ID**
  - Uniquely identifies the publisher's account within the ad system hosted by the company
- **Seller type**
  - DIRECT or RESELLER
- **Certification Authority ID** (TAG)
  - Optional
  - an ID that uniquely corresponds to the company in field one

Data collection

- Our focus: News sources
- Sampling frame: List of URLs
- Sampling frame sourced from 3 directories:
    - Jasmine directory
    - W3newspapers
    - USNPL
- Representation
    - Representative of all 50+1 states
- Limitations
    - Data source creation methodology
   - Partial compliance of ads.txt standard

Data cleaning

1. Base URL extraction
   ○ "https://example.com/somepage" to "example.com"
2. Data enrichment
   ○ Enriching state information and topic information wherever available
3. Deduplication
   ○ Deduping if the same url exists in multiple directories
4. Valid URLs extraction
   ○ Does "[base url]/ads.txt" give an HTTP 200 response?
   ○ Accounting for redirects
      i. If there is a redirect to a different base URL then mapping created.
      ii. Valid URL extraction done on redirects
5. Ads.txt file extraction
   ○ Accounting for bad files such as placeholder files
6. Edge list creation
   ○ Combining files
   ○ Removing improper lines in the files

Creation of graph DB ([Neo4j](Neo4j))

- A simple data model with two entities (node types) and a uni-directional edge
  - [:Publishers]-[:USES]->[:Platforms]
- 3 files were created:
  - Platforms.csv assigns each platform a unique key in addition to its URL
  - Publishers.csv assigns each platform a unique key in addition to its URL
  - Uses.csv maps platform key to publisher key along with a relationship type attribute and number of relationship count attribute
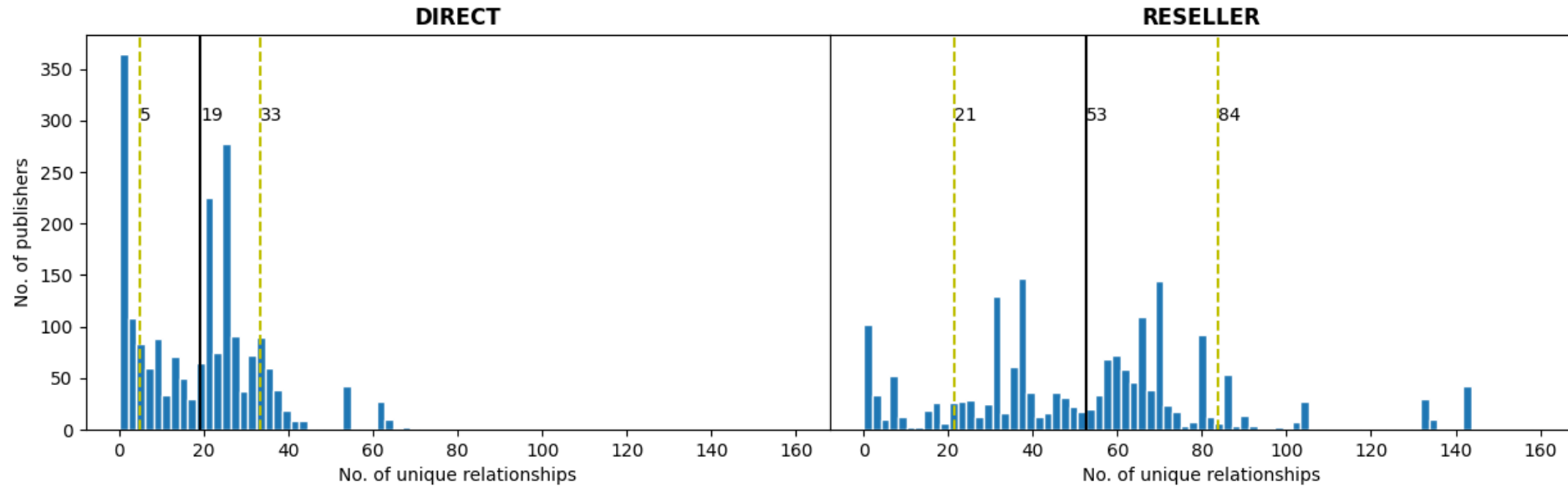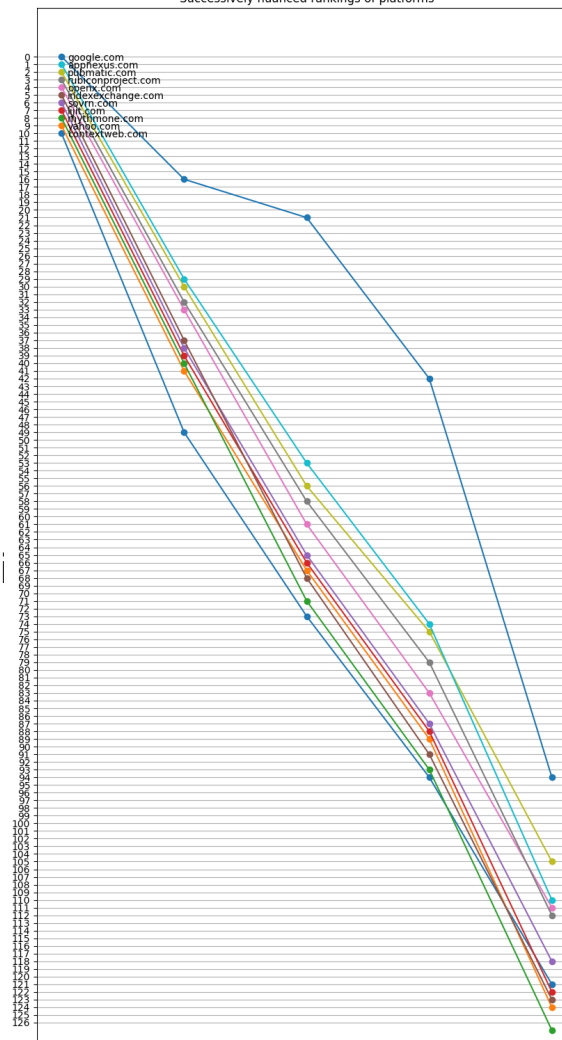
Descriptive analysis

Counts:

Publishers 2,050

U
S
E
S

Unique: 137,038

Non unique: 1,064,527
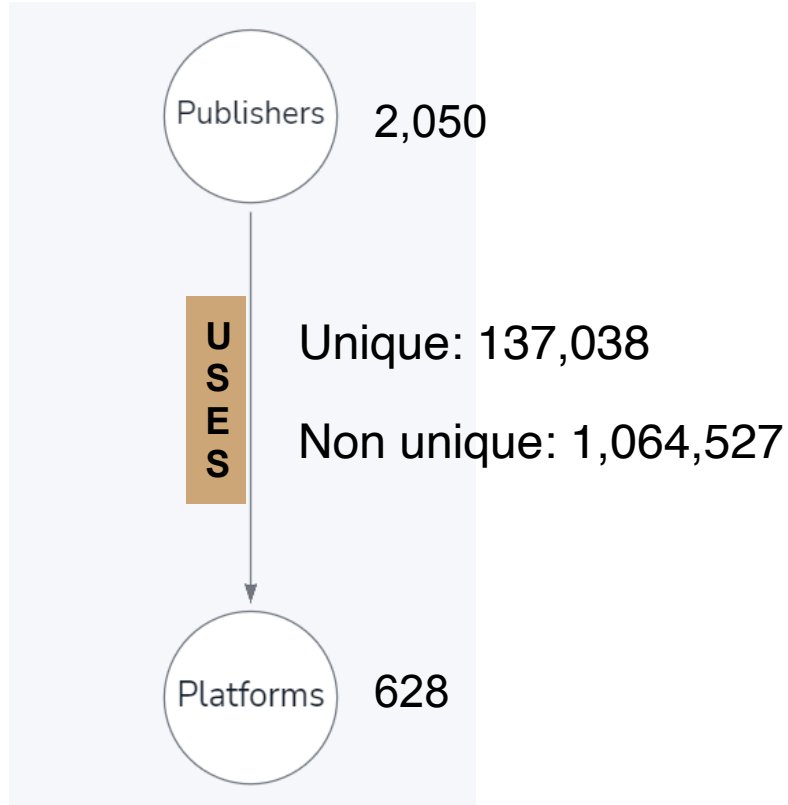
Platforms 628

Descriptive analysis

Platforms matter!

Successively nuanced rankings of platforms

Bipartite network analysis
- Method of reflections

- Observe that adding more nuance brings the rankings down substantially
- Are current definitions of market concentration limiting?

Conclusion

- Our contributions:
  - We have added empirical evidence to structural theory
  - We have provided analytical methodologies for studying this data
- How things designed better in future
  - Expand the graph model
    - Include sellers.json
  - Transition to big data
    - Scrap a lot more publishers (not just newspapers)
  - Studying the data over time
    - Using "Wayback machine"
  - Developing better metrics for studying concentration and overall market complexity.