comscore

Graphing Crumbling Cookies

AdKDD 2019

Matt Malloy, Jon Koller and Aaron Cahn

What is a device graph?

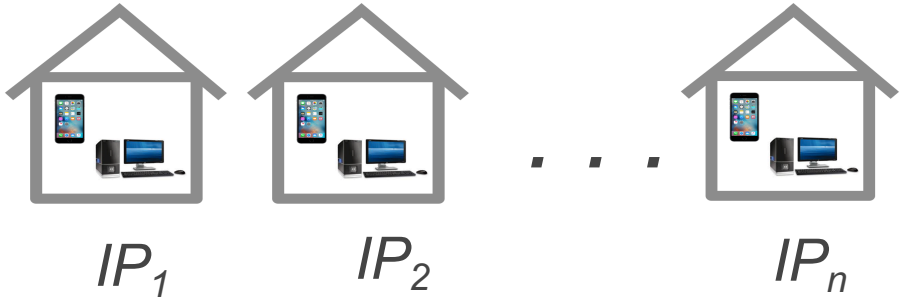
- a dataset that organizes digital identifiers that we create as we use the internet
- identifiers (IDs): browser cookies or advertising IDs
- a *graph* is a set of vertices and edges
- a list of pairs of identifiers that are in some way related

id_1	id_2	score
3D0F8F	*54D3A8*	3.936
7F3E10	*6FFE0A*	1.400
8764CF	*10AFC8*	3.440
501EE5	*62A1F3*	3.045
1F39D3	*4B2686*	4.763
638581	*85B16*	1.917

- related: same person, same household
- example: *two digital IDs that login with same email*
- *Why?* ***Targeting, content customization and accurate measurement***

Building a graph using IP-colocation

- IP space is ***intimate***
  - Your devices share an IP when connected to the same WiFi router
  - You share an IP with family, friends and co-workers
- ideal world: static residential IPs



- problem: IPs are dynamic, mobile operator/corporate IPs, coffee shops
- observation: even when IP changes, devices travel through IP-space together over course of weeks

basic idea: associate devices with each other, not IP

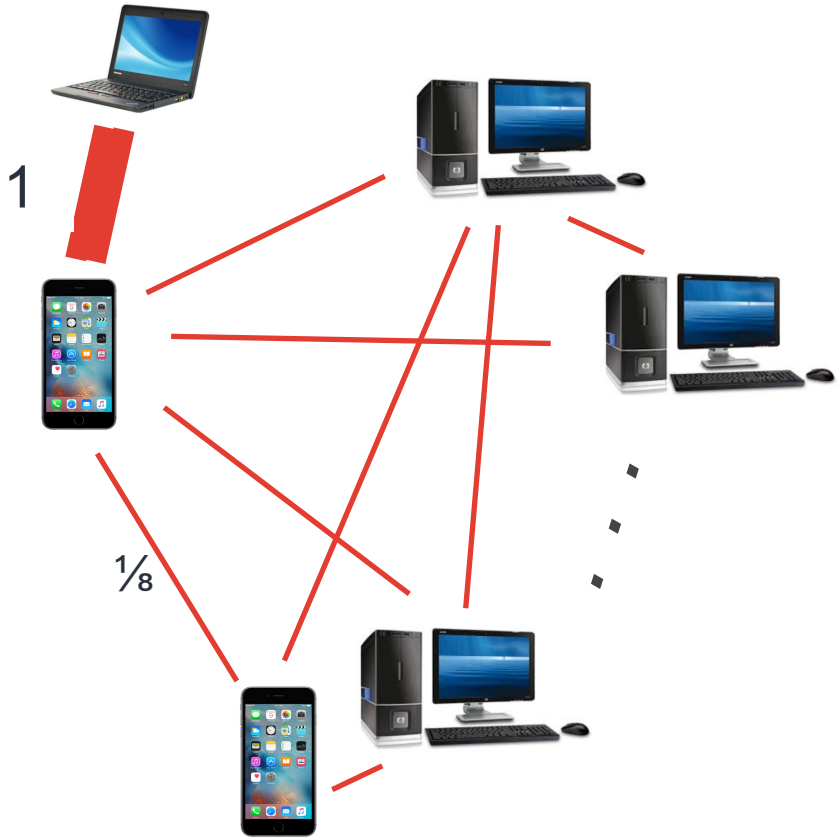
Building a graph

day 1: iPhone is home with PC

day 2: iPhone is home alone

day 3: iPhone is at work with 8 devices

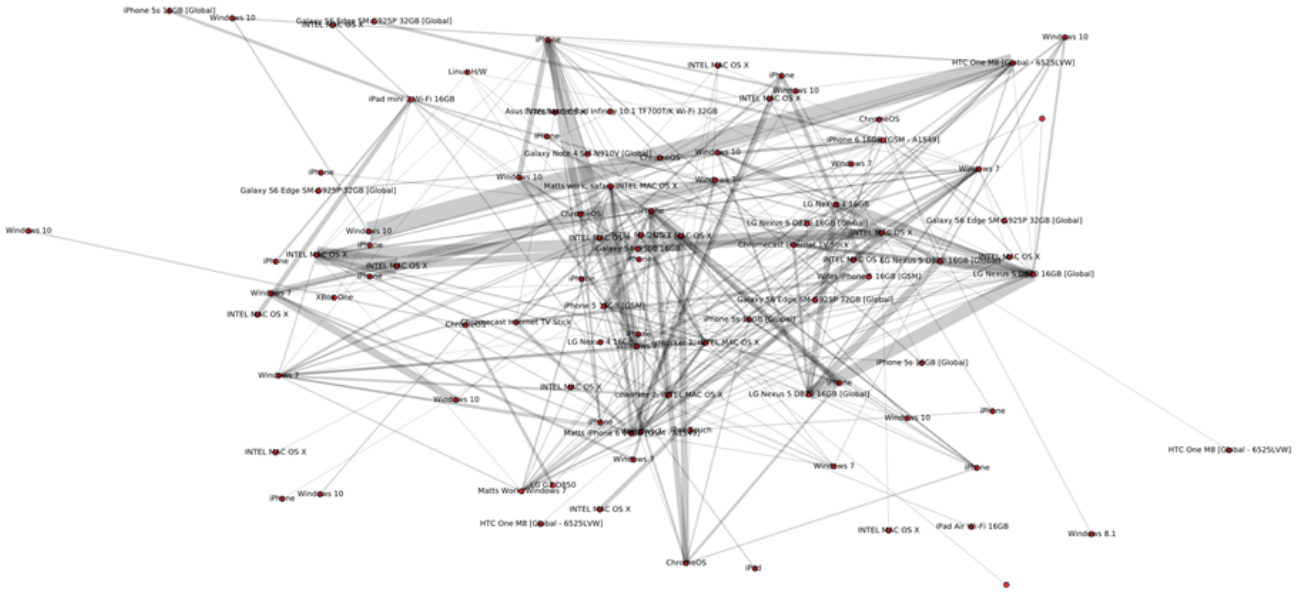
day 4: iPhone is at home with PC



- score proportional to number of days two devices spend alone on an IP

Malloy, M., Barford, P., Alp, E. C., Koller, J., & Jewell, A. (2017, August). Internet Device Graphs. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1913-1921). ACM.

Comscore's Device Graph



Comscore's Device Graph (April 2019)

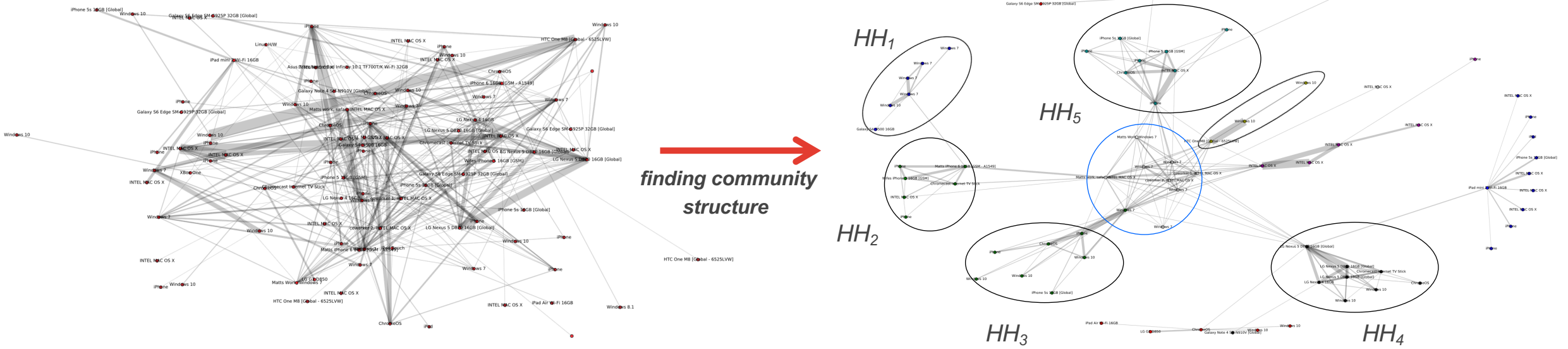
- 12 countries
- 3.4 Billion nodes (cookies/advertising IDs)
- 17.1 Billion edges (relationships)

Comparison Benchmark Graphs*

Graph	Nodes	Edges
LiveJournal	4.8M	69M
Twitter	42M	1.5B
UK web graph 2007	109M	3.7B
Yahoo Web	1.4B	6.6B
Facebook Graph 2016	1.39B	400B

*Adapted from: Ching, A., Edunov, S., Kabiljo, M., Logothetis, D., & Muthukrishnan, S. (2015). One trillion edges: Graph processing at facebook-scale. Proceedings of the VLDB Endowment, 8(12), 1804-1815..

Community Detection



- goal: group identifiers into cohorts (person and household level groupings)
- community detection in graphs is a well studied problem
  - Literature/code for finding community structure (but not billions of nodes/edges)
  - Louvain Modularity*

*Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

Challenge: non-persistent IDs

- 3.4 Billion **persistent** IDs (in 12 countries)

- 5-10x more **non**-persistent IDs
  - excluded from graphing process
  - incognito/private browsing (session cookies)
  - ITP (Intelligent Tracking Prevention)

- 20+ Billion IDs worldwide not amenable to graphing or community detection

Backfilling

Key Ideas:

- Once cohorts of persistent IDs are defined, find the IP addresses that are associated with the cohort over time:

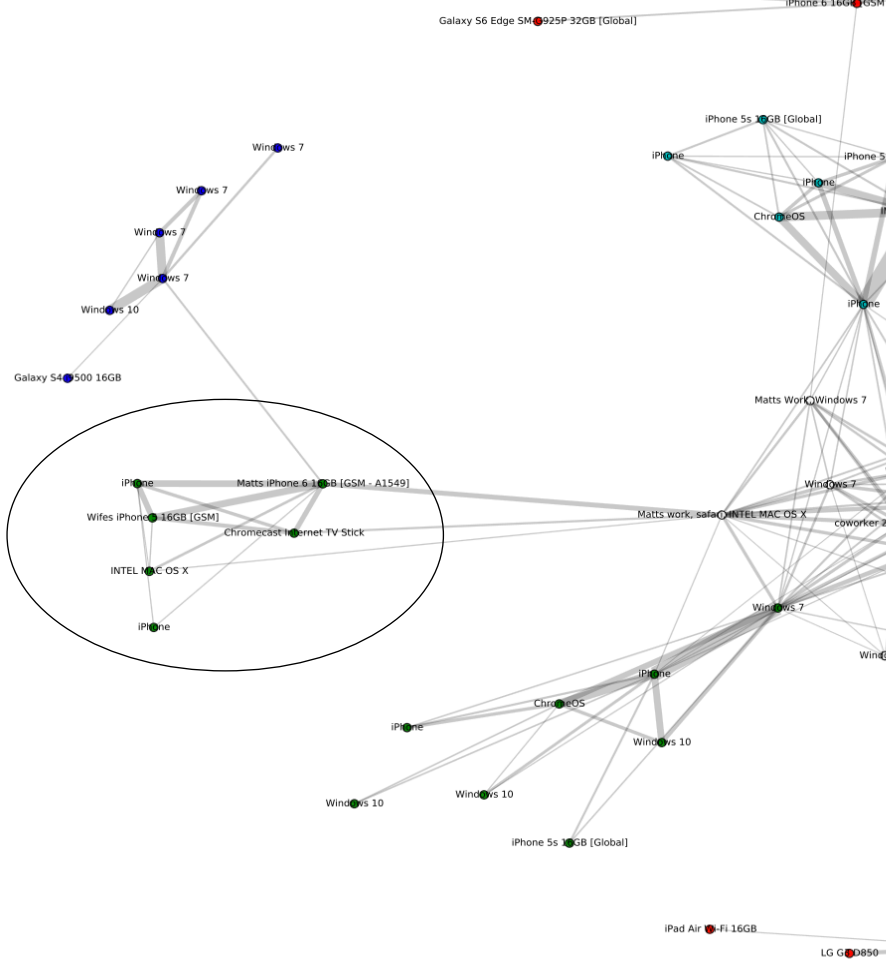
$$\mathcal{C}_1 \rightarrow \{(\mathrm{IP}_1, \mathrm{day}_1), (\mathrm{IP}_2, \mathrm{day}_2), \dots\}$$

- Ruleset: if the persistent IDs defined by the IP address are synonymous with the group defined by cohort, then assign non-persistent IDs to cohort:

$$\text{if } \{i : i \in (\mathrm{IP}_1, \mathrm{day}_1)\} \cap V_\mathrm{p} \approx \mathcal{C}_1$$
$$\text{then } C_1^+ = \{i : i \in (\mathrm{IP}_1, \mathrm{day}_1)\} \cup C_1$$

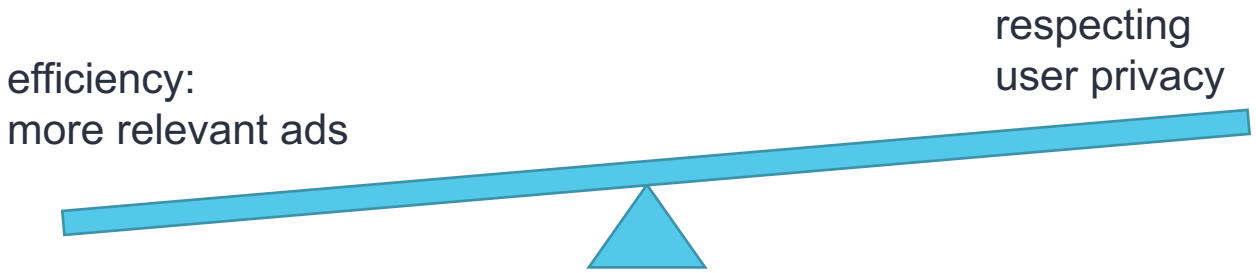
- Precision and recall are used to define approximate equality ($\approx$)

- Results: assign additional 2+ Billion IDs to cohorts in the US



Privacy

- Internet is great. It's funded by ads.

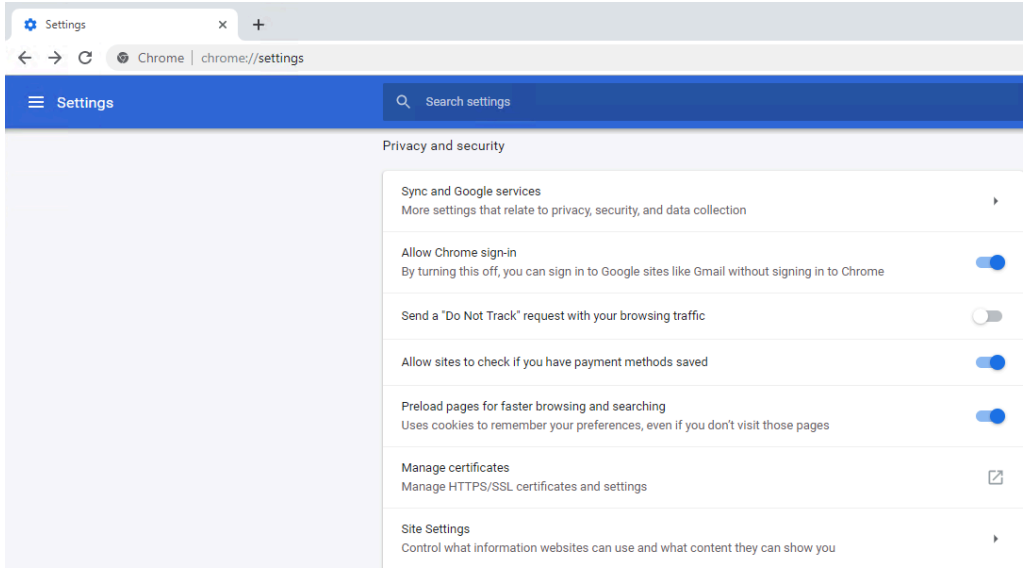


- Current/future landscape
  - Increases in non-persistent identifiers and rejection of 3rd party cookies
  - Safari, Firefox, likely more to come
  - Legislation - GDPR (Europe) and CCPA (California)

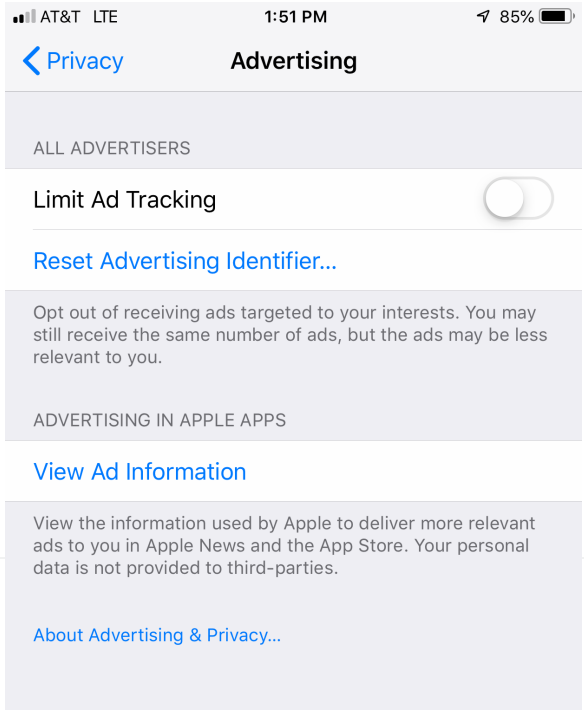
- Favor large entities with login information (Google, Facebook, Apple)

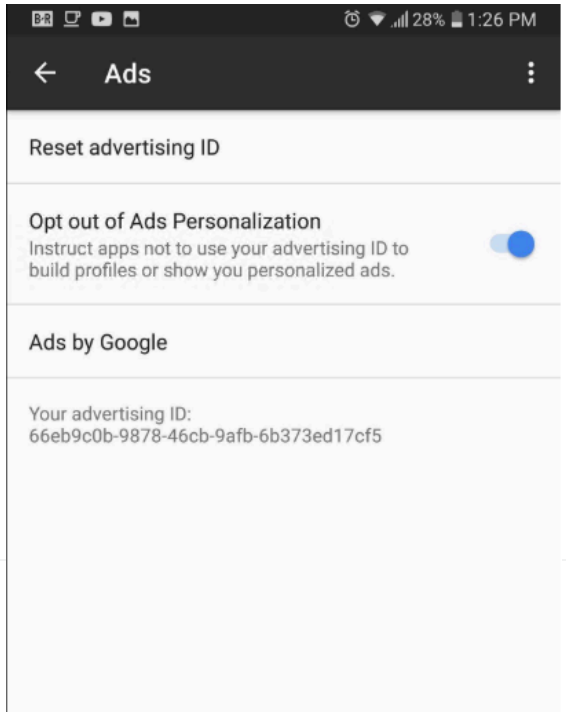
How to opt-out

- Reject 3rd party cookies.



- Turn off your advertising ID.





Questions?

Device Graph Publications

- Graphing Crumbling Cookies, AdKDD (Malloy, Koller, Cahn)
- Device Graphing by Example, KDD 2018 (Funkhouser, Malloy, Alp, Poon, Barford)
- Internet Device Graphs, KDD 2017 (Malloy, Barford, Alp, Koller, Jewell)