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.

<table>
<thead>
<tr>
<th>id_1</th>
<th>id_2</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D0F8F</td>
<td>54D3A8</td>
<td>3.936</td>
</tr>
<tr>
<td>7F3E10</td>
<td>6FFE0A</td>
<td>1.400</td>
</tr>
<tr>
<td>B764CF</td>
<td>10AFC8</td>
<td>3.440</td>
</tr>
<tr>
<td>501EE5</td>
<td>62A1F3</td>
<td>3.045</td>
</tr>
<tr>
<td>1F39D3</td>
<td>4B2686</td>
<td>4.763</td>
</tr>
<tr>
<td>638581</td>
<td>85B16</td>
<td>1.917</td>
</tr>
</tbody>
</table>

- 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
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*

<table>
<thead>
<tr>
<th>Graph</th>
<th>Nodes</th>
<th>Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiveJournal</td>
<td>4.8M</td>
<td>69M</td>
</tr>
<tr>
<td>Twitter</td>
<td>42M</td>
<td>1.5B</td>
</tr>
<tr>
<td>UK web graph 2007</td>
<td>109M</td>
<td>3.7B</td>
</tr>
<tr>
<td>Yahoo Web</td>
<td>1.4B</td>
<td>6.6B</td>
</tr>
<tr>
<td>Facebook Graph 2016</td>
<td>1.39B</td>
<td>400B</td>
</tr>
</tbody>
</table>

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*

---

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:
  \[ C_1 \rightarrow \{(IP_1, day_1), (IP_2, day_2), \ldots \} \]

- 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:
  \[
  \text{if } \{i : i \in (IP_1, day_1)\} \cap V_p \approx C_1 \\
  \text{then } C_1^+ = \{i : i \in (IP_1, day_1)\} \cup C_1
  \]

- 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)