Taming the long tail
Identify Filtering in Social Media

Avner May, Nitish Korula, Silvio Lattanzi
A. Chaintreau
Columbia University, Google
joint work with Avner, Nitish and Silvio
Are social media sustainable?
From the trenches: no!

- **Users’s dilemma**
  - May I be missing something?

- **Content producters**
  - May I be missing my audience?
From the faculty lounge: of course!

- **Socializing** is essential for information
  - To find about jobs [Gr74], innovation [CKM57]
    “It pays to know / It hurts to be unaware.”

- When looking for good content, most of the time is wasted, but some gems are priceless
  - This process is more efficient collectively
  - And curating is at least informally rewarded

- In this talk, we focus on **news** dissemination
What is the role of intermediaries?
Understanding these intermediaries
2013: two interesting works

Twitter “precision” 40.5% average - Encouraging!

Homogeneous or structured interests leads to efficient networks
Can we find evidence of filtering?
Looking for filtering

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Source</th>
<th>Users</th>
<th>URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY Times Links</td>
<td>Twitter</td>
<td>330k</td>
<td>33k</td>
</tr>
<tr>
<td>Bin Laden Death</td>
<td>Twitter</td>
<td>700k</td>
<td>545k</td>
</tr>
<tr>
<td>Occupy Wall Street</td>
<td>Twitter</td>
<td>354k</td>
<td>316k</td>
</tr>
<tr>
<td>Steve Jobs Death</td>
<td>Twitter</td>
<td>719k</td>
<td>251k</td>
</tr>
<tr>
<td>iPhone 5 Launch</td>
<td>Twitter</td>
<td>81k</td>
<td>37k</td>
</tr>
<tr>
<td>iPhone 5 Launch</td>
<td>Facebook</td>
<td>330k</td>
<td>193k</td>
</tr>
<tr>
<td>All Spinn3r blogs</td>
<td>Spinn3r</td>
<td>68k</td>
<td>441k</td>
</tr>
<tr>
<td>Obama</td>
<td>Spinn3r</td>
<td>13k</td>
<td>85k</td>
</tr>
<tr>
<td>Facebook</td>
<td>Spinn3r</td>
<td>12k</td>
<td>70k</td>
</tr>
<tr>
<td>Euro</td>
<td>Spinn3r</td>
<td>10k</td>
<td>53k</td>
</tr>
<tr>
<td>Mubarak</td>
<td>Spinn3r</td>
<td>7k</td>
<td>43k</td>
</tr>
</tbody>
</table>
Evidence of information filtering

“Filtering law”

Not an artefact of
  - replacement
  - exposure

MORE ACTIVITY → LESS POPULAR CONTENT
Many open questions

- Can we find more evidence of precision?
  - Using click (Twitter data grant, more partners)
  - Does selectivity correlate with success?

- Current models somewhat at odds
  - Discrete topics + continuous popularity range
  - Are there more general models

- Can crowd-curation be improved?
  - In principle (no friction etc.), already efficient.
  - With incentive? With new mechanism?
Thank you!
Back-Up Slides
# Theoretical Results

<table>
<thead>
<tr>
<th>Audience Strategy</th>
<th>Pure Strategy Equilibrium?</th>
<th>Price of Anarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greedy</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>Satisficing</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>Satisficing w/ blogger ability</td>
<td>Yes</td>
<td>2</td>
</tr>
</tbody>
</table>
Filtering Law Consistent Across Data Sets
INACTIVE
< 2 / month
5%

ACTIVE
< 2 / day
35%

VERY ACTIVE
>= 2 / day
60%

MORE ACTIVE CONTENT ➔ LESS POPULAR
Simply explained by replacement effect? **NO!**
In Summary…

- Previous work: Intermediaries play key role in information dissemination.
- We provided theoretical and empirical justification for intermediaries as information filters.
- Come see my poster!
  - Results not shown: Role of filtering on success of intermediary