Copyright, Digitization, and Aggregation

Lesley Chiou and Catherine Tucker

Occidental College and MIT Sloan

April 2012
Google News is typical of aggregators.
Aggregation is controversial
Aggregation is controversial

- Online aggregators assert that they only display extracts of information, which is often factual.
Aggregation is controversial

- Online aggregators assert that they only display extracts of information, which is often factual.

- Content providers fear that consumers may use extracts as substitutes for accessing and reading the full content. Aggregators are “content kleptomaniacs” (Diskin, 2011).
How are consumers using aggregators?

- Do consumers use aggregators to lower transactions costs and terminate search for content that they would already seek?
How are consumers using aggregators?

- Do consumers use aggregators to lower transactions costs and terminate search for content that they would already seek?
- Do consumers seek new content that they would not otherwise obtain? If so, what types of content benefit?
We study a discontinuous change in provision of content
We study a discontinuous change in provision of content

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 2006</td>
<td>Google and The Associated Press first sign contract (30 day window)</td>
</tr>
<tr>
<td>December 24, 2009</td>
<td>The Associated Press content no longer appears on Google.</td>
</tr>
<tr>
<td>End January 2010</td>
<td>The Associated Press and Google contract set to expire</td>
</tr>
<tr>
<td>February 2010</td>
<td>The Associated Press content returns to Google News.</td>
</tr>
</tbody>
</table>
Michigan voters search for economic savior

By KATHY BARKS HOPKINSN (AP) — 1 day ago

LANSLNG, Mich. — Michigan voters frustrated over lost jobs, home foreclosures and budget deficits will vote in Tuesday’s primary election for leaders they hope can move the state out of its economic morass.

With seven men running for governor and nearly two dozen candidates running for three open congressional seats, the hardest task may be sorting through the barrage of names, campaign ads and economic rhetoric.

The candidates and voters agree that Michigan is at a crossroads. After a decade of malaise that has left the state with the nation’s second-highest unemployment rate and one in every four residents relying on unemployment insurance, Medicaid, cash assistance or food stamps, creating more jobs is the overwhelming priority and topic of debate.

The gubernatorial candidates are competing to succeed outgoing Democratic Gov. Jennifer Granholm, who can’t run again because of term limits and whose popularity sank with her struggles to revive the economy.

All seven gubernatorial candidates say they plan to cut business taxes to attract employers. Most of the five Republicans also say they would slash state regulations and cut state spending. One, Oakland County Sheriff Mike Bouchard, proposes getting rid of laws forcing workers to join unions to get certain jobs.

Among the Democrats, Lansing Mayor Virg Bernero is visiting factory gates and union halls to pledge he’ll stand up for middle-class workers and jobs. His opponent, Andy Dillon, a business turnaround specialist who’s now the House speaker, promises to bring in more alternative energy jobs to replace lost manufacturing work.

With platforms that are similar, the Republicans are using their job credentials to assure voters they would be the best at managing the economy.

National GOP interest in unseating freshmen Democratic Reps. Mark Schauer in mid-Michigan’s 7th District and Gary Peters in the Detroit suburbs in Oakland County has Republicans vying in both districts for the chance at a November matchup.

Copyright © 2016 The Associated Press. All rights reserved.

Related articles
Gov candidates woo undecided voters
during final weekend before primary
The Detroit News - 7 hours ago

Race to get names on November ballot in
Michigan governor’s race is wide open
Lansing State Journal - 16 hours ago

Editorial: Focus on jobs, not abortion in
governor race
The Detroit News - 21 hours ago

More coverage (1)
Users do not view aggregation as a substitute for copyrighted content.

Traffic to news sites after visiting Google News experience 0.6 ppt drop in share of clicks after the removal of AP articles (compared to Yahoo! News).
Users do not view aggregation as a substitute for copyrighted content.

Traffic to news sites after visiting Google News experience 0.6 ppt drop in share of clicks after the removal of AP articles (compared to Yahoo! News).

Aggregation benefits content that is horizontally or vertically differentiated.

Removal of content led to decline in visits to either very regionally concentrated or national websites.
We measure how an expansion or contraction of content affects navigation from the platform.
We measure how an expansion or contraction of content affects navigation from the platform

- We do not have the type of individual-level data that would allow us to study direct navigation to aggregators vs. other content providers
We measure how an expansion or contraction of content affects navigation from the platform

- We do not have the type of individual-level data that would allow us to study direct navigation to aggregators vs. other content providers (which is why we are happy that Athey and Mobius have a new paper that tackles this!)
We measure how an expansion or contraction of content affects navigation from the platform

- We do not have the type of individual-level data that would allow us to study direct navigation to aggregators vs. other content providers (which is why we are happy that Athey and Mobius have a new paper that tackles this!)

- Conditional on the existence of aggregators, how does the addition or removal of content affect subsequent searches from the platform?
Our research has policy implications for intellectual property rights and copyright

No empirical work has quantified the effect of aggregators on primary sources

- US Copyright Office requires proof of “fair use”
- Does the use of the copyrighted material reduce its potential market or value?
Our research has policy implications for intellectual property rights and copyright

No empirical work has quantified the effect of aggregators on primary sources

- US Copyright Office requires proof of “fair use”
- Does the use of the copyrighted material reduce its potential market or value?

Intellectual property and copyright enforcement on the Internet

- Do aggregators violate existing copyrights?
- Should content be provided freely by a third party?
Outline

1. Introduction
2. Data
3. Analysis
4. Implications
Data

Experian Hitwise

- Internet Service Providers (ISP)
- 25 million people worldwide
- Websites visited by users after navigating to Google News or Yahoo! News
- Fraction of total traffic at each “downstream” site
- Weekly data from December 2009 to January 2010
Types of websites

News sites—e.g., nytimes.com
Types of websites

News sites—e.g., nytimes.com

Non-News sites
- International—e.g., hindustantimes.com
- Top aggregators
- Weather
- Other—e.g., hotmail.com
Summary statistics

Table: Summary statistics for downstream websites from Google News and Yahoo! News

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>% clicks</td>
<td>0.016</td>
<td>0.19</td>
<td>0</td>
<td>18.3</td>
<td>98730</td>
</tr>
<tr>
<td>Google News</td>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>Yahoo! News</td>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>PeriodDispute</td>
<td>0.67</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>News Site</td>
<td>0.20</td>
<td>0.40</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>Non-news Site</td>
<td>0.80</td>
<td>0.40</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>Aggregator Site</td>
<td>0.00091</td>
<td>0.030</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>International Site</td>
<td>0.048</td>
<td>0.21</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>Weather Site</td>
<td>0.0067</td>
<td>0.081</td>
<td>0</td>
<td>1</td>
<td>98730</td>
</tr>
<tr>
<td>Observations</td>
<td>98730</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Users of Yahoo! News vs. Google News

Table: Demographic description of users

<table>
<thead>
<tr>
<th>Measure</th>
<th>Yahoo! News</th>
<th>Google News</th>
<th>New York Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>59.95</td>
<td>63.8</td>
<td>61.21</td>
</tr>
<tr>
<td>Age 18-24</td>
<td>12.12</td>
<td>13.89</td>
<td>6.17</td>
</tr>
<tr>
<td>Age 25-34</td>
<td>18.05</td>
<td>14.72</td>
<td>13.93</td>
</tr>
<tr>
<td>Age 35-44</td>
<td>19.03</td>
<td>17.08</td>
<td>12.98</td>
</tr>
<tr>
<td>Age 45-54</td>
<td>21.41</td>
<td>22.24</td>
<td>19.45</td>
</tr>
<tr>
<td>Age 55+</td>
<td>29.38</td>
<td>32.06</td>
<td>47.47</td>
</tr>
<tr>
<td>Income &lt;30k</td>
<td>22.33</td>
<td>20.77</td>
<td>20.76</td>
</tr>
<tr>
<td>Income 30-60k</td>
<td>28.82</td>
<td>27.53</td>
<td>26.36</td>
</tr>
<tr>
<td>Income 60-100k</td>
<td>24.95</td>
<td>24.6</td>
<td>24.82</td>
</tr>
<tr>
<td>Income 100-150k</td>
<td>14.61</td>
<td>17.5</td>
<td>17.29</td>
</tr>
<tr>
<td>Income &gt;150k</td>
<td>9.29</td>
<td>9.6</td>
<td>10.77</td>
</tr>
</tbody>
</table>

Source: Hitwise
No global changes occurred in usage of Yahoo! News and Google News during this period, Yahoo! News and Google News experienced no changes in overall rankings in traffic, average visit time, and number of pages navigated.
Pre- and post- AP removal from Google News

Average percentage of downstream visits to news sites falls after AP removal from Google News
December 2008 and January 2009

No change in average percentage of downstream traffic between December 2008 and January 2009
Empirical strategy

For website $i$ after visiting news aggregator $j$ in week $t$:

\[
\% \text{clicks}_{ijt} = \beta_0 + \beta_1 \text{News}_i \times \text{Google}_j \times \text{PeriodDispute}_t \\
+ \beta_2 \text{News}_i \times \text{PeriodDispute}_t \\
+ \beta_3 \text{News}_i \times \text{Google}_j + \beta_4 \text{Google}_j \\
+ \alpha_i + \text{week}_t + \epsilon_{ijt}
\]

where

- $\text{News} = 1$ if website is a news source
- $\text{Google} = 1$ if the traffic originated after viewing Google News
- $\text{PeriodDispute} = 1$ for the weeks after the removal of Associated Press from Google News
Table: Downstream traffic from Google and Yahoo! News before and after the policy change

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeriodDispute × Google × News</td>
<td>-0.00600***</td>
<td>-0.00599***</td>
<td>-0.00623***</td>
<td>-0.00622***</td>
</tr>
<tr>
<td></td>
<td>(0.00289)</td>
<td>(0.00289)</td>
<td>(0.00305)</td>
<td>(0.00305)</td>
</tr>
<tr>
<td>PeriodDispute × Google</td>
<td>0.00159</td>
<td>0.00158</td>
<td>0.00182</td>
<td>0.00181</td>
</tr>
<tr>
<td></td>
<td>(0.00233)</td>
<td>(0.00234)</td>
<td>(0.00253)</td>
<td>(0.00253)</td>
</tr>
<tr>
<td>Google</td>
<td>-0.0119*</td>
<td>-0.0117*</td>
<td>-0.0156**</td>
<td>-0.0155**</td>
</tr>
<tr>
<td></td>
<td>(0.00638)</td>
<td>(0.00640)</td>
<td>(0.00696)</td>
<td>(0.00698)</td>
</tr>
<tr>
<td>PeriodDispute × News</td>
<td>0.00135</td>
<td>0.00134</td>
<td>0.00137</td>
<td>0.00136</td>
</tr>
<tr>
<td></td>
<td>(0.00102)</td>
<td>(0.00102)</td>
<td>(0.00104)</td>
<td>(0.00104)</td>
</tr>
<tr>
<td>News × Google</td>
<td>0.0326***</td>
<td>0.0324***</td>
<td>0.0363***</td>
<td>0.0361***</td>
</tr>
<tr>
<td></td>
<td>(0.00775)</td>
<td>(0.00777)</td>
<td>(0.00823)</td>
<td>(0.00825)</td>
</tr>
<tr>
<td>Website Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>98730</td>
<td>98640</td>
<td>93951</td>
<td>93861</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.581</td>
<td>0.580</td>
<td>0.581</td>
<td>0.580</td>
</tr>
</tbody>
</table>
Back-of-the-envelope calculations

News sites on Google experience 0.6 ppt decrease in clicks after the removal of The Associated Press articles

- 20% decrease in visits
Back-of-the-envelope calculations

News sites on Google experience 0.6 ppt decrease in clicks after the removal of The Associated Press articles

- 20% decrease in visits
- 1 billion clicks to Google news partners in a month
Back-of-the-envelope calculations

News sites on Google experience 0.6 ppt decrease in clicks after the removal of The Associated Press articles

- 20% decrease in visits
- 1 billion clicks to Google news partners in a month
- Domestic sites: 40%
Back-of-the-envelope calculations

News sites on Google experience 0.6 ppt decrease in clicks after the removal of The Associated Press articles

- 20% decrease in visits
- 1 billion clicks to Google news partners in a month
- Domestic sites: 40%

Overall: 80 million decrease in clicks
How does aggregation affect different types of content?
How does aggregation affect different types of content?
Outline

1. Introduction
2. Data
3. Analysis
4. Implications
Conclusion

Aggregation inspires people to seek new content (of a more unusual and high quality).
Conclusion

Aggregation inspires people to seek new content (of a more unusual and high quality).

* Traffic to news sites after visiting Google News experience 0.6 ppt drop in clicks after the removal of AP articles (compared to Yahoo! News).
Conclusion

Aggregation inspires people to seek new content (of a more unusual and high quality).

- Traffic to news sites after visiting Google News experience 0.6 ppt drop in clicks after the removal of AP articles (compared to Yahoo! News).

- Sites with content that is either horizontally or vertically differentiated may benefit from aggregation.
Our results have implications for the “Information Economy” and public policy

- Digital advances reduce search costs and promote greater search for information
  - “Hyper-local” sites and “microcontent”
  - Survival of “quality” content in the Information Age
Our results have implications for the “Information Economy” and public policy

- Digital advances reduce search costs and promote greater search for information
  - “Hyper-local” sites and “microcontent”
  - Survival of “quality” content in the Information Age

- US Copyright Office: 4-part test

- Digital Millenium Copyright Act and copyright reform