The Web and its Publics

Tommaso Venturini

In the talk, I will discuss the social and political consequences of the organization of digital media. I will consider the limits of a simplistic reading of the power-law distribution of online visibility and the hopes raised by the thematic clustering and the dynamism of the Web. I will also touch on the risks that these dynamics entail exploring the causes of the recent proliferation of 'junk news'.

La démocratie Internet

Dominique Cardon
Seuil (2010)

On a beaucoup souligné les origines militaires d’Internet... Mais la chose est désormais établie. Internet est surtout né de la rencontre entre la contre-culture américaine et l’esprit méritocratique du monde de la recherche (p. 1).

Chercheurs, artistes, militants, passionnés et freaks de toutes espèces, les premier publics de l’Internet y ont fait proliférer des utopies futuristes, des expérimentations esthétiques, des provocations et des gestes politiques d’un nouveau genre (p. 2).

The Internet Imaginaire

Patrice Flichy
2007, MIT Press

The founding utopias of computer communication not only guided the initial Arpanet project but also constantly interacted with its technical realization (p. 65).

[For hackers the use and future of IT] was not only an intellectual tool for academics but also a device to put into everyone’s hands, capable of building not only new invisible colleges but also a new society (p. 67).
The first many-to-many medium of communication

A new Athens
Albert Gore
Remarks at International Telecom Union
Buenos Aires, 21 March 1994

“In a sense, the Global Information Infrastructure will be a metaphor for democracy itself... it will in fact promote the functioning of democracy by greatly enhancing the participation of citizens in decision-making. And it will greatly promote the ability of nations to cooperate with each other. I see a new Athenian Age of democracy forged in the fora the Global Information Infrastructure will create.”

Digital divide

Infrastructure bottlenecks
Internet censorship

The rise of platforms

Internet traffic to and from Egypt on January 27 - 28. At 5:20 pm EST, traffic to and from Egypt across 83 internet providers around the world drops precipitously.

And even if we forget about all this

Does technical symmetry guarantees communication equality?

Irving Kristol (1920-2009):
“Democracy does not guarantee equality of conditions it only guarantees equality of opportunities”
The Web as a scale-free network


The probability of finding very popular addresses, to which a large number of other documents point, is non-negligible, an indication of the flocking sociology of the www (p. 72).

The law of power

The Web as a narrowcasting medium


When the power to filter is unlimited, people can decide, in advance and with perfect accuracy, what they will and will not encounter (p. 5).

I will emphasize the risks posed by any situation in which thousands or perhaps millions or even tens of millions of people are mainly listening to louder echoes of their own voices. A situation of this kind is likely to produce far worse than mere fragmentation (p. 16).
The Web as a broadcasting medium

Democracy?
Not at the top nor at the bottom

The Web concentration

<table>
<thead>
<tr>
<th>Three Metrics of Media Concentration</th>
<th>Gin coeff.</th>
<th>HHI</th>
<th>Norn index</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Web sites</td>
<td>.76</td>
<td>.69</td>
<td>22</td>
</tr>
<tr>
<td>News and media sites</td>
<td>.88</td>
<td>.78</td>
<td>40</td>
</tr>
<tr>
<td>Political sites</td>
<td>.85</td>
<td>.80</td>
<td>31</td>
</tr>
<tr>
<td>Radio audience</td>
<td>.53</td>
<td>.19</td>
<td>—</td>
</tr>
<tr>
<td>Newspaper circulation</td>
<td>.69</td>
<td>.73</td>
<td>18</td>
</tr>
<tr>
<td>Magazine circulation</td>
<td>.70</td>
<td>.12</td>
<td>34</td>
</tr>
</tbody>
</table>

Metrics of Concentration for Newspapers

<table>
<thead>
<tr>
<th>Top 10</th>
<th>Gin</th>
<th>HHI</th>
<th>Norn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers—Print circulation (top 200)</td>
<td>.30%</td>
<td>.50</td>
<td>143</td>
</tr>
<tr>
<td>Newspapers—Web site visits (top 200)</td>
<td>.42%</td>
<td>.62</td>
<td>304</td>
</tr>
</tbody>
</table>

What about the middle?
To some extent we've always consumed media that appealed to our interests and avocations and ignored much of the rest. But the filter bubble introduces three dynamics we've never dealt with before.

First, you're alone in it. A cable channel that caters to a narrow interest (say, golf) has other viewers with whom you share a frame of reference. But you're the only person in your bubble. In an age when shared information is the bedrock of shared experience, the filter bubble is a centrifugal force, pulling us apart.

A newly created webpage will more likely include links to well known, popular documents with already high connectivity (p. 73)

For unto every one that have shall be given, and he shall have abundance: but from him that have not shall be taken away even that which he have» (p. 59)

A scientific contribution will have greater visibility in the community of scientists when it is introduced by a scientist of high rank (p. 60)

Confronted with the growing task of identifying significant work published in their field, scientists search for cues to what they should attend to. One such clue is the professional reputation of the authors (p. 60)
Two difficulties in methodology

1. **Semantic**
   hyperlinks are embedded in texts that influence the decision to follow a link

2. **Time**
   hyperlinks are created successively and more recent links are more relevant

The missing middle?

Matthew Hindman
The Myth of Digital Democracy

<table>
<thead>
<tr>
<th>Audience Share for Online and Off-line Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>All Web sites</td>
</tr>
<tr>
<td>News and media</td>
</tr>
<tr>
<td>Political sites</td>
</tr>
<tr>
<td>Radio audience</td>
</tr>
<tr>
<td>Newspaper circulation</td>
</tr>
<tr>
<td>Magazine circulation</td>
</tr>
</tbody>
</table>

Composition of Audience Share, Print versus Digital

<table>
<thead>
<tr>
<th>Top 10</th>
<th>11-20</th>
<th>21-50</th>
<th>51-100</th>
<th>101-500</th>
<th>501+</th>
</tr>
</thead>
<tbody>
<tr>
<td>News and media Web sites</td>
<td>29%</td>
<td>12%</td>
<td>10%</td>
<td>9%</td>
<td>21%</td>
</tr>
<tr>
<td>vs. newspaper circulation</td>
<td>-10%</td>
<td>-12%</td>
<td>-7%</td>
<td>-4%</td>
<td>-7%</td>
</tr>
<tr>
<td>vs. magazine circulation</td>
<td>+2%</td>
<td>+3%</td>
<td>-6%</td>
<td>-8%</td>
<td>-20%</td>
</tr>
</tbody>
</table>

4 stirring mechanisms

From the Bottom to the middle
1. (S) The Web is not just clustered, it is fractally clustered
2. (T) Ranking within clusters can change rapidly

From to top to the middle
3. (S) Winner takes all, but different clusters have different winners
4. (T) Stable distribution but unstable composition

The Web visibility stream

- In-Degree (received citations)
- Thematic prominence
- Fractal clustering
- Stable composition
- N° of websites
“We are not intellectual lemmings”

It turns out that we are not intellectual lemmings. We do not use the freedom that the network has made possible to plunge into the abyss of incoherent babble.

Instead, through iterative processes of cooperative filtering and “transmission” through the high visibility nodes, the low-end thin tail turns out to be a peer-produced filter and transmission medium for a vastly larger number of speakers than was imaginable in the mass-media model (p. 255)

The “Fake News” shock

https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.qnG0mJyJb#vixod5y58

and yet…

"Fake News" and the Public Data Lab

fakenews.publicdatalab.org

publicdatalab.org

"We are not intellectual lemmings”

Yochai Benkler
The Wealth of Networks
Yale University Press (2006)
according to the literature, the notion of fake news is

1. Vague
2. Undistinguishable from past forms of misinformation
3. Charged with a simplistic idea of journalistic truth
4. Politically dangerous

4 reasons to stay clear of “fake news”
The term "fake news" has also begun to be appropriated by politicians around the world to describe news organisations whose coverage they find disagreeable. In this way, it’s becoming a mechanism by which the powerful can clamp down upon, restrict, undermine and circumvent the free press. It’s also worth noting that the term and its visual derivatives (e.g., the red ‘FAKE’ stamp) have been even more widely appropriated by websites, organisations and political figures identified as untrustworthy by fact-checkers to undermine opposing reporting and news organizations” (p.16)
Economy (automatization of the online attention market)
Communication (socialization of a ‘prosumer’ audience)
Technology (behavioral algorithms and spreading bots)
Culture (virality-oriented subcultures)
Politics (trolling warfare)

“Macron is gay” hoax

Harmful news bloom
Google’s success came from an understanding of what Chris Anderson refers to as “the long tail,” the collective power of the small sites that make up the bulk of the web’s content. The Web 2.0 lesson: leverage customer self-service and algorithmic data management to reach out to the entire web, to the edges and not just the center, to the long tail and not just the head (p.21).

The hit economy

Using politics to influence virality (not the other way around)

"The young Macedonians who run these sites say they don’t care about Donald Trump. They are responding to straightforward economic incentives... Several teens and young men who run these sites told BuzzFeed News that they learned the best way to generate traffic is to get their politics stories to spread on Facebook — and the best way to generate shares on Facebook is to publish sensationalist and often false content that caters to Trump supporters."


Detecting fake news is not the name of the game

"When looking at just the titles of fake and real news articles, I find an even stronger dissimilarity between the two, with high consistency between the data sets and high statistical significance in the differences. Precisely, I find that fake news titles are longer than real news titles and contain simpler words in both length and technicality. Fake titles also used more all capitalized words, significantly more proper nouns, but fewer nouns overall, and fewer stop-words... When adding in satire articles to the analysis, I find that the majority of our features distributions are common between satire and fake. Specifically, both satire and fake use smaller, fewer technical, and fewer analytic words, as well as, fewer quotes, fewer punctuation, more adverbs, and fewer nouns than real articles. Further, fake and satire use significantly more lexical redundancy..."


Debunking fake news is not the name of the game

"I can’t help but laugh at the irony of folks screaming up and down about fake news and pointing to the story about how the Pope backs Trump. The reason so many progressives know this story is because it was spread wildly among liberal circles who were citing it as appalling and fake. From what I can gather, it seems as though liberals were far more likely to spread this story than conservatives. What more could you want if you ran a fake news site whose goal was to make money by getting people to spread misinformation? Getting doubters to click on clickbait is far more profitable than getting believers because they’re far more likely to spread the content in an effort to dispel the content. Win-a"a

The raise of click baiting and native advertising

"Real-time, auction-specific quality calculations of expected clickthrough rate, ad relevance, and landing page experience, among other factors, are used to calculate Ad Rank at auction time. These factors, which are based on things known only at the time of the auction, can heavily influence the quality of the user’s experience."


You don’t even have to read to take the bait
(Data brokers & third party cookies)

AdSense revenue share

For displaying ads with AdSense for content, publishers receive 68% of the revenue recognized by Google in connection with the service. For AdSense for search, publishers receive 75% of the revenue recognized by Google. These percentages are constant, regardless of a publisher’s geographic location, and are not in any way affected by the publisher. We don’t charge the revenue share for other AdSense products; the revenue share varies for other products due to different costs of developing and supporting these products.

You can view revenue share details in your account:
1. Sign in to your AdSense account.
2. In the left navigation panel, click Settings.
3. Click Account, then click Account Information.
4. In the “Account information” section, you’ll see each revenue share displayed next to “Active products.”

Active products
- AdSense: 68% publisher revenue share
- AdSense for search: 75% publisher revenue share
- Mobile: 75% publisher revenue share
- Video: 75% publisher revenue share

We believe our revenue share is extremely competitive. However, revenue shares alone can be misleading, so we encourage you to focus on the total revenue generated for your site. For example, if Google’s auction of an impression on your site generates $1.10, with our 68% revenue share you’d receive $0.69 through AdSense. Another ad network might offer an 80% revenue share, but only collect $0.65 from advertisers, so you’d receive $0.52.

With the vast number of advertisers competing to appear on AdSense sites, our system ensures that you’re earning the most possible for every ad impression you receive.
Behavioral recommendation algorithms

Covington, Paul, Jay Adams, and Emre Sargin. 2016
Deep Neural Networks for YouTube Recommendations
Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16

The rise of social buttons

1% rule of Internet culture

90% Lurkers

Posters and lurkers

And of social metrics
How to Define Your Personal Brand

What are your values and how do you want to get these across?
What tone of voice would you like to use?
Which is your personal audience and what kind of audience would you like to target?
What do you want to achieve with your brand?
What are you, and what is your position in the industry?

Personal branding and micro-celebrities

How 4chan promotes viral ephemerality

How 4chan promotes viral ephemerality

Pizzagate and LOLcats

4chan and virality oriented subcultures

Political astroturfing


“the team doing this research... fail[s] to understand the legitimacy of “public opinion guidance” within the Chinese system... traditional public opinion guidance systems don’t seem to be pulling their weight when it comes to overcoming these problems. The Internet media space has an infinite capacity but its borders and its core are unclear, so some grassroots social issues are always able to suddenly attract the attention of the entire Internet, creating one hot button issue after another”