Opinion Dynamics via Search Engines
(and other algorithmic gatekeepers)

Fabrizio Germano
UPF & Barcelona GSE

Francesco Sobbrio
LUISS "G. Carli"

AGCOM
March 17, 2017

Motivation: Algorithmic Gatekeepers

- Search engines and social media use algorithms to *rank info*.

- Examples of ranked info via automated algorithms:
  - Search results obtained from a search query in *Google* or *Bing*.
  - News/tweets observed in *Facebook* or *Twitter*.

- Ranking algorithms determine *what info is relevant* for an individual.
  - Search engines & social media as "`algorithmic gatekeepers`"
    (Introna, Nissenbaum 2000, Granka 2010, Pariser 2011, Tufekci 2015)

This paper: What are the implications for opinion dynamics?

- Main focus: **search engines**.

This Paper

We develop a **theoretical framework** where:

- Individuals use a search engine to look for info.
- Stylized structure of a search engine's algorithm.
- Algorithm interacts with individuals' online search behavior.

We then study the following issues:

- Evolution of website rankings
- Individuals' choices over websites
- Individuals' beliefs over given issues (*opinion dynamics*)
- Efficiency and asymptotic learning.

Main Assumptions: Search Engine's Ranking Algorithm

Ranking parameters are grouped into three main categories:

1. *Ex-ante* parameters on website "authority" (e.g., indexing, PageRank)
   - initial ranking
2. Usage data (e.g., website clicks, *Facebook* likes, *Twitter* links)
   - dynamic component *(popularity)*
3. Parameters to personalize search results (e.g., users' IP's).
   - personalized vs. non-personalized ranking algorithm.

Main Assumptions: Individuals' Search Behavior

Two main assumptions on individuals' use of search engines:

1) Search cost $\Rightarrow$ Higher ranked websites more likely to be chosen

- As in search diversion lit (Hagiu & Jullien, 2014)

- Empirical (causal) evidence of individuals more likely to choose higher-ranked websites (Glick et al. 2011; Pan et al., 2011; Epstein & Robertson, 2015).

2) Preference for like-minded news $\Rightarrow$ Like-minded websites more likely to be chosen

- As in media econ lit (Mullainathan & Shleifer, 2005, Gentzkow, Shapiro, 2010)

- Empirical (causal) evidence that individuals are more likely to click on like-minded websites (Flaxman et al. 2013; Yom-Tov et al. 2013)

Theoretical Framework

Information structure:

- Binary state of the world, $\left(\frac{1}{2}, \frac{1}{2}\right)$ Bernoulli random variable, $\omega \in \{0, 1\}$

M websites:

- Receive a private signal with accuracy $q > \frac{1}{2}$
- Simply report their own private signal (non-strategic).

N individuals:

- Need to choose an action to match ω (e.g., whether to vaccinate child)
- Receive a private signal with accuracy $p > \frac{1}{2}$
- $q > p$, i.e., websites have higher accuracy than individuals.
- Individuals (sequentially) use search engine to look for info on ω

Search engine:

- Provides a ranking of websites to each individual *snippets*
- Updates the ranking based on websites' *popularity*

Incentive Compatible Benchmark

No search cost & no preference for like-minded news

Since $q > p$:

- *Ex-ante* most informative signal:
  - "website-majority" signal (e.g., mainstream news websites).
- Each individual:
  - rationally "ignores" her private signal
  - chooses any of the K websites reporting the website-majority signal.
- Probability of individual n choosing website m:
  - $\rho_{n,m} = 1/K$
  - ranking does not matter.

Full Model

Search cost ($\alpha > 0$) & preference for like-minded news ($\gamma > 0$)

- Probability of individual n choosing website m:

$$\rho_{n,m} = f(r_{n,m}, \alpha, \gamma)$$

  - $r_{n,m}$: ranking of m at time n
    ($\uparrow r_{n,m}$: more likely to choose m)
  - α: search cost
    ($\uparrow \alpha$: more likely to choose higher ranked website)
  - γ: preferences for like-minded news
    ($\uparrow \gamma$: more likely to choose website confirming prior).

Search Engine & Ranking Algorithm

- Universal ranking algorithm $\mathcal{R}$:

$$r_{n,m} = (1-\nu) r_{n-1,m} + \nu \rho_{n-1,m},$$

  - The lower ν, the more persistent the ranking is.

- Personalized ranking algorithm $\mathcal{R}_\ell$:

$$r^\ell_{n,m} = \nu^\ell_n \rho_{n-1,m} + (1-\nu^\ell_n) r^\ell_{n-1,m},\ \ell = A, B,$$

  - 2 Groups: $A, B \subset N$, such that $A \cup B = N$ and $A \cap B = \emptyset$.
  - $\nu^\ell_n = \nu^\ell_n(\lambda)$ depends on whether $n-1$ is in the same group as n.
  - λ: extent of search result personalization
    (i.e., higher λ, lower weight on choices of individuals from $\neq$ group).

Example

At time $t = n$:

- A parent gets private signal regarding side effects of vaccines
- Uses search engine to get other info
- Observes the ranking of the websites $r_{n,m}$
- Given ranking, search cost α, and preferences for like-minded websites γ: decides which website to read
- The ranking algorithm updates the ranking of websites according to:

$$r_{n+1,m} = (1 - \nu) r_{n,m} + \nu \rho_{n,m}$$

At time $t = n + 1$....

Overview of the Main Results

1. Popularity ranking + search cost $\Rightarrow$ *rich get richer*
2. Personalization + preferences for like-minded news $\Rightarrow$ *belief polarization*

$\Rightarrow$ High market concentration & Ideological Segregation

3. *Advantage of the fewer*: fewer websites with incorrect info each attract more traffic $\Rightarrow$ gain a higher rank $\Rightarrow$ attract even more traffic.

$\Rightarrow$ Rationale of why "fake news" may thrive with algorithmic gatekeepers

4. Efficiency and Asymptotic Learning $\Rightarrow$ Personalization:
   1. Sub-optimal on common value issues (e.g., side effects of a vaccine)
   2. Useful on private value issues (e.g., attributes of a commercial product).

Extensions

- Domain Bias ($M' \subset M$): main dynamics still apply.
- Non random search order: path dependence

Policy Implications

Our framework $\Rightarrow$ No incentives to strategically manipulate search results:

- To increase their advertising profits (as in De Corniere & Taylor 2014; Burguet et al. 2015; Hagiu and Jullien 2014)
- To influence election outcomes (as in Epstein & Robertson, 2015)

Yet, even in this "best scenario "$\Rightarrow$ subtle distortions

Rationale for public intervention:

- *Rich-get-richer* $\Rightarrow$ competition authorities
- *Belief polarization* $\Rightarrow$ communication authorities