

Filter bubbles and algorithmic filter bubbles

Matthew J. Salganik

Social Network (Soc 204)
Spring 2017
Princeton University

March 29, 2017



Logistics:

- ▶ there are no wrong answers to survey questions
- ▶ precept this week

Fake News and Informed Citizens

“Fake News on Facebook is all the rage right now, but what do we know about its impacts and how it spreads? Join CITP Fellow Katherine Haenschen for a discussion of recent research pertaining to Fake News and how information spreads on Facebook generally.”

Katherine Haenschen Thu, 03/30/2017 - 12:15pm 011 Robertson Hall
RSVP: <http://lapa.princeton.edu/content/fake-news-and-informed-citizens>

1. Pariser, E. (2011). Beware online 'filter bubbles'. TED talk.
2. Weisberg, J. (2011). Bubble Trouble: Is Web personalization turning us into solipsistic twits. Slate.
3. Bakshy, E., Messing, S., and Adamic, L.A. (2015) Exposure to ideologically diverse news and opinion on Facebook. Science.
4. Sandvig, C. (2015) The Facebook 'it's not our fault' study. Social Media Collective Research Blog.



https://commons.wikimedia.org/wiki/File:Reflection_in_a_soap_bubble_edit.jpg

Today we are going to talk about filter bubbles in Facebook and Google.

Disclosure: I have received research funding from Facebook, Google, and Microsoft.

[https://theintercept.com/2016/09/15/
tech-money-lurks-behind-government-privacy-conference/](https://theintercept.com/2016/09/15/tech-money-lurks-behind-government-privacy-conference/)

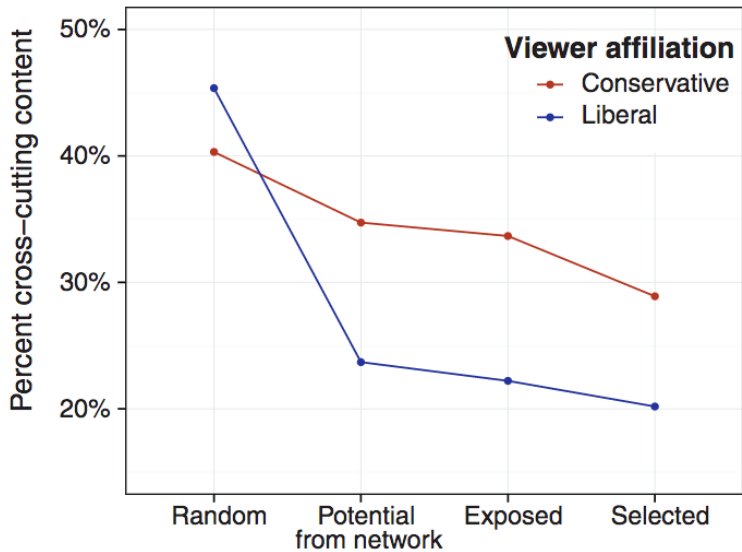
Exposure to ideologically diverse news and opinion on Facebook

Eytan Bakshy,^{1*†} Solomon Messing,^{1†} Lada A. Adamic^{1,2}

Information exposure is a multi-step process

Information exposure is a multi-step process

1. people you are connected to
2. stories that your feed prioritizes
3. stories that you click on



Who is included?

- ▶ “we utilize a large, comprehensive dataset from Facebook.”
- ▶ “we examined how 10.1 million U.S. Facebook users interact”
- ▶ in appendix: “we consider active U.S. adults on Facebook who report their political affiliation.”

Who is included?

- ▶ “we utilize a large, comprehensive dataset from Facebook.”
- ▶ “we examined how 10.1 million U.S. Facebook users interact”
- ▶ in appendix: “we consider active U.S. adults on Facebook who report their political affiliation.”
- ▶ “18 or older”
- ▶ “log in at least 4/7 days per week”
- ▶ “have interacted with at least one link shared on Facebook that we classified as hard news”
- ▶ “self-report their ideological affiliation in a way that was interpretable”

Who is included?

- ▶ “we utilize a large, comprehensive dataset from Facebook.”
- ▶ “we examined how 10.1 million U.S. Facebook users interact”
- ▶ in appendix: “we consider active U.S. adults on Facebook who report their political affiliation.”
- ▶ “18 or older”
- ▶ “log in at least 4/7 days per week”
- ▶ “have interacted with at least one link shared on Facebook that we classified as hard news”
- ▶ “self-report their ideological affiliation in a way that was interpretable”

Does this sampling matter?

How did they measure the ideology of the subject?

How did they measure the ideology of the subject?

All Facebook users can self-report their political affiliation; 9% of U.S. users over 18 do. We mapped the top 500 political designations on a five-point, -2 (Very Liberal) to +2 (Very Conservative) ideological scale; those with no response or with responses such as “other” or “I don’t care” were not included. 46% of those who entered their political affiliation on their profiles had a response that could be mapped to this scale. We validated a sample of these labels against a survey of 79 thousand U.S. users in which we asked for a 5-point very-liberal to very-conservative ideological affiliation; the Spearman rank correlation between the survey responses and our labels was 0.78.

How did they measure the ideology of the subject?

All Facebook users can self-report their political affiliation; 9% of U.S. users over 18 do. We mapped the top 500 political designations on a five-point, -2 (Very Liberal) to +2 (Very Conservative) ideological scale; those with no response or with responses such as “other” or “I don’t care” were not included. 46% of those who entered their political affiliation on their profiles had a response that could be mapped to this scale. We validated a sample of these labels against a survey of 79 thousand U.S. users in which we asked for a 5-point very-liberal to very-conservative ideological affiliation; the Spearman rank correlation between the survey responses and our labels was 0.78.

Does this sampling matter?

How did they measure the ideology of the post?

How did they measure the ideology of the post?

“We derive the alignment score A of an individual URL by averaging the political alignment of the set of people who share the URL.”

How did they measure the ideology of the post?

“We derive the alignment score A of an individual URL by averaging the political alignment of the set of people who share the URL.”

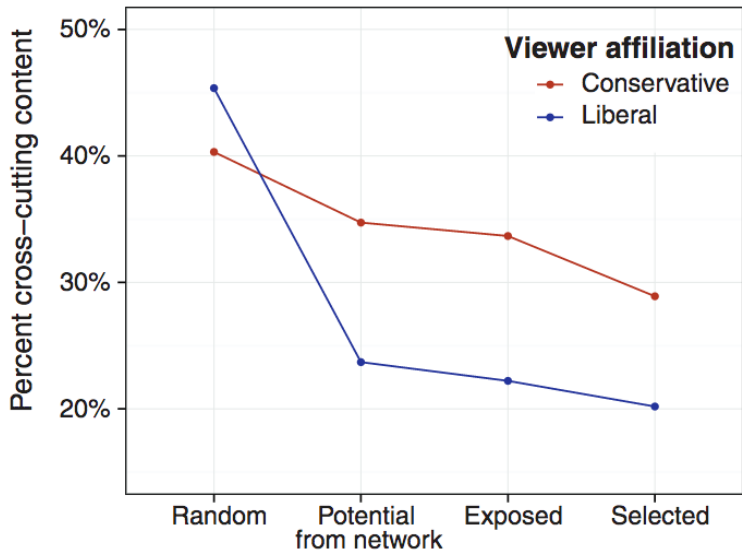
Does this coding scheme matter?

What kind of posts?

What kind of posts?

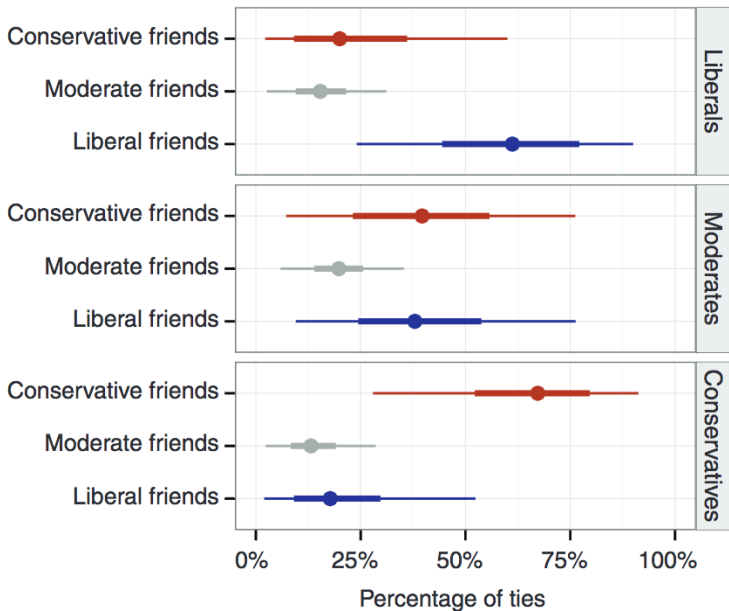
“We classified stories as either “hard” (such as national news, politics, or world affairs) or “soft” content (such as sports, entertainment, or travel) by training a support vector machine on unigram, bigram, and trigram text features (details are available in the supplementary materials, section S1.4.1).”

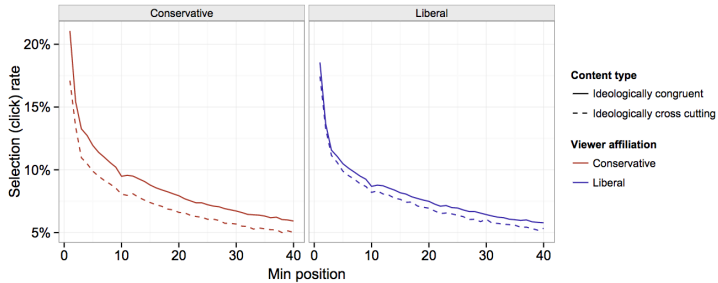
Approximately 13% were classified as hard news. Or approximately 87% were classified as soft news (not counting baby pictures).



Two interesting, non-filter bubble findings

F





Descriptive vs normative

Is it Facebook's fault?

- ▶ “Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content.”
- ▶ “Within the population under study here, individual choices more than algorithms limit exposure to attitude-challenging content in the context of Facebook.”

Is it Facebook's fault?

- ▶ “Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content.”
- ▶ “Within the population under study here, individual choices more than algorithms limit exposure to attitude-challenging content in the context of Facebook.”

What should Facebook do?

Additional commentary and criticism:

- ▶ [https://medium.com/message/
how-facebook-s-algorithm-suppresses-content-diversity-](https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-)
- ▶ [https://thesocietypages.org/cyborgology/2015/05/
07/facebook-fair-and-balanced/](https://thesocietypages.org/cyborgology/2015/05/07/facebook-fair-and-balanced/)
- ▶ [http://crookedtimber.org/2015/05/07/
why-doesnt-science-publish-important-methods-info-prom](http://crookedtimber.org/2015/05/07/why-doesnt-science-publish-important-methods-info-prom)

Could someone that didn't work at Facebook do this research?

Could someone that didn't work at Facebook do this research?
Does that matter?

The Filter Bubble

What [redacted] the [redacted]
[redacted]
[redacted] Internet [redacted]
[redacted]
[redacted] Is [redacted]
[redacted]
[redacted] Hiding [redacted]
[redacted]
[redacted] From [redacted]
[redacted]
[redacted] You [redacted]

Eli Pariser

Pariser:

- ▶ Egypt
- ▶ BP

Pariser:

- ▶ Egypt
- ▶ BP

Weisberg (name sound familiar?):

- ▶ John Boehner
- ▶ Barney Frank
- ▶ Ryan Plan
- ▶ Obamacare

Pariser:

- ▶ Egypt
- ▶ BP

Weisberg (name sound familiar?):

- ▶ John Boehner
- ▶ Barney Frank
- ▶ Ryan Plan
- ▶ Obamacare

Us: ?

What terms should we use?

What terms should we use?

What types of terms would we expect a lot of personalization between members of this class?

- ▶ discussion about terms
- ▶ wiki survey
- ▶ everyone does search and screenshots, upload to piazza
- ▶ start analyzing data

Summary:

- ▶ Algorithmic filter bubbles are only one part of a multi-step process that impacts information exposure

Summary:

- ▶ Algorithmic filter bubbles are only one part of a multi-step process that impacts information exposure
- ▶ It is not clear that algorithmic filter bubbles are the most important part of that system

Summary:

- ▶ Algorithmic filter bubbles are only one part of a multi-step process that impacts information exposure
- ▶ It is not clear that algorithmic filter bubbles are the most important part of that system
- ▶ Algorithmic filter bubbles are very hard to measure for a single person

To be continued in precept. . . .

<http://bit.ly/socnet204>

<http://bit.ly/socnet204>

- ▶ Explore Blue Feed, Red Feed, FlipFeed, and/or Escape Your Bubble.
- ▶ Eslami, et al. (2015) "I always assumed that I wasn't really that close to [her]:" Reasoning about invisible algorithms in the news feed. Proceedings of the 33rd Annual SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery (ACM): 153-162. (and summary video).
- ▶ Constine, J. (2016) How Facebook News Feed Works.
- ▶ Roberts, S. (2001) Surprises from self-experimentation: sleep, mood, and weight. Chance.
- ▶ Rubin, D.B. (2001) Comment: self-experimentation for causal effects. Chance.