ALGO 119 - Module 3_1

Hello and welcome back to week 3.

Today we're going to talk about the role platform algorithms play in curating online news content, but before we get into that I want to remind you that there's another video lecture this week as well as one longer reading on algorithmic news distribution that you should also have a look at. The longer reading is actually a chapter from my book, so I hope you really enjoy it. Also don't forget to engage in the forums this week, and to take the online quiz.

So for the rest of this video today I want to talk about how algorithms run by platforms like Google, Facebook and Apple are really powerful in terms of directing attention to news information online.

These algorithms make important curation decisions about which content to show and make more prominent, which can have profound implications based on the scale in which they operate online.

Facebook and Google are really the 800 pound gorilla's here. They run some of the most powerful and important curation algorithms in the world which impact what billions of people see and interact with on a daily basis. So a few statistics here, there are 45 percent of U.S. adults that get news on Facebook, more than half of U.S. adults get news from Google, and about a quarter of people around the world say that Google is the main gateway to news information for them.

Facebook and Google also account for about 70 percent of the external traffic driven to news sites. So that last statistic, actually comes from the parsley media dashboard which you can see here, and you can see the dominance of Google and Facebook here in terms of their share of external referrals to traffic.

This chart shows the external referrals to news sites in the parceling network over the past 12 months roughly, and you can see Google is at the top with 48 percent, Facebook is second with 27 percent, and then way down in third place is Flipboard, but that's just two point four percent. So really you see that Google and Facebook totally dominate the external referral traffic to news sites.

So one of the reasons that news curation at the scale of Google and Facebook is so important and can matter so much is that it has the potential to impact important things like voting turnout. So consider the results of an internal study done at Facebook in 2012 shown in this slide here.

In this study the data scientists at Facebook looked at the impact of showing people more hard news information in their newsfeeds, so they compared the regular version of the news feed curation algorithm labeled as control in this slide to an experimental version of the algorithm labeled as boost in this slide.

In the boosted experimental condition they turned up the visibility of hard news information that was seen by people in the news feed. So you can see in both the boost condition as well as the control condition there was an increase in voter turnout that was measured.

In the top panel they're comparing for people who log in less frequently, and in the bottom panel they're comparing information for people who log in every day to the site, but in either case there's an increase in voter turnout.

So this study really shows just the totally awesome power of news curation algorithms to impact democratic processes like voting, and It should really make us all I think a little bit wary of the power that algorithms like Facebook wield in society.

Now it's no surprise that Facebook plays with its curation algorithm fairly frequently. In January 28 they posted this blog post introducing a change to how they were prioritizing posts in the news feed algorithm. In this case they change the algorithm so that privileged posts that generate more conversation, so more comments around the post.

This meant that posts that generate a lot of discussion like say political controversies will get even more exposure in the news feed, whereas other posts that maybe are on topics that are more personal or maybe even taboo to discuss, those posts because they're not generating discussion would get less exposure from the news feed algorithm.

So that's why every little tweak and change that's made to a curation algorithm has to be sort of inspected and critiqued to understand its real implications. Besides Facebook, Google is another online platform that we all probably use on an almost daily basis, and there's a tremendous amount of power embedded in how Google directs attention to information including news information online.

About a third of news content online is found by a search, and again more than half of Americans get news via search. So if search for a news topic on Google you'll often get a set of results at the top labeled as in the news for top stories.

So not only does Google integrate news into its regular search results, but it also draws special attention to a set of news articles that their algorithm curates for this top stories section. So, in my research lab we actually got really curious about this and in particular about which sources were popping up in this top stories box.

And we gathered some data to look at this, and look at the diversity of sources that were showing up in this top stories box, and this is what we found. So, just looking at one day's worth of news on one search, we gathered all the news articles in this top articles pop box for a search on tax reform.

This was a topic that was trending and was in the news in the U.S. at that time, and every minute for 24 hours we scraped the articles that were showing up on Google, and what we found was quite surprising actually. There were only 12 different news sources that showed up in that entire day, and if you look at these you can see Washington Post there is at the top followed by CNBC and Politico.

Moving down the list there's really only one of those sources that you might consider to be a local source, that's Buffalo News with three percent of the articles that showed up there, and some other experiments

that we ran during the 2016 elections, we actually found that four searches for Hillary Clinton and for Donald Trump, we found that there were just two news sources The New York Times and CNN which dominated about 44 percent of the news links that we found.

So again, this is just one example, it's just for one query on one particular day, but what we found is that there wasn't a lot of diversity in the news sources that Google was exposing searchers to.

And that can have all kinds of implications in terms of the diversity of perspectives that people are exposed to, as well as the amount of traffic that's being driven to different news outlets. In this case, more traffic being driven to say large news outlets like The Washington Post than to more local news outlets.

Another type of algorithmic curation algorithm that we often encounter online are trends algorithms, and those typically look at the popularity of a hashtag or an event, and then put it into some highly visible list which then amplifies that trend even more because more and more people are seeing it.

So similar to our interest in the diversity of news sources that we found on google, we were also curious to see what kinds of sources were found in Apple news. In particular in the Apple News trending section, which is an algorithmically curated list of the trending stories.

Now for this story, for this experiment, we just ran an initial pilot study and we collected data once every hour for about three days by asking people online to take screenshots of their Apple news trends and then we were able to tabulate the sources that were popping up based on the screenshots that people sent into us.

So here are the top seven sources that we've found, and in total there were only 25 unique sources that we observed over this almost three day period. Now you can see that Fox News sort of tops the ranking there followed by People and CNN, Huffington Post, Vanity Fair, Bleacher Report, and The New York Times.

I think what's interesting here is that while some of these news sources do focus on on news, several of them are also more about celebrity and entertainment. For instance People and Vanity Fair, and that sort of gives you a flavor for the composition of hard news versus soft news that Apple is curating in its trending news section of its app.

So to recap here today, I'd like to focus on a couple of key takeaway points for you. The first is that curation decisions can lead to all kinds of consequential impacts for society. They can impact public opinion, they can impact and influence democratic processes like voting and voter turnout, they have implications for whether people are able to find each other and form communities online, and the impact the diversity of information exposure that people have online.

The other point to drive home is how important these platforms for algorithms are for driving traffic to news content. This can have implications for smaller publishers as well as larger publishers that maybe aren't getting as much of a boost in traffic as they maybe could be from these platform algorithms.

And because of the power of platforms and driving exposure to news information, media publishers typically have to try to orient to the metrics and the algorithms of the platforms in order to succeed.

I'll talk about this more in the next video in terms of how publishers are optimizing content for news distribution, so I'll see you there.