First Response Essay:
The Production of Information in an Online World

March 29, 2023

1 Research Question & Motivation

In their paper *The Production of INformation in an Online World* the authors Julia Cagé, Nicolas Hervé and Marie-Luce Viaud research the following question:

> **Research Question**
> Given the limited intellectual property protection in news media, what is the extent of copying in online news production, and what are the incentives to produce original content?

While their motivation is complex, we can boild it down to two main concerns:

1. Influence of bad journalism on the political landscape.

2. How to safeguard the quality of journalism.

Both those concern come from the point of view, that press and news media are the so called fourth estate which basically is the believe, that press and news media have a responsibility of providing quality journalism to properly inform the voters. Furthermore, they argue that current copyright laws are too weak and need to be stronger, to prevent an "abuse of power" from the fourth estate.

2   Contributions

Comparing to previous work, the authors have to following contributions:

- Data: Construction of a new, very detailed and large data set.

- Methodology: Application of new and state-of-the-art NLP algorithm.

- Methodology: Creation of a new plagiarism algorithm.

Furthermore, they state that their work also complements a growing empirical literature on copyright.

3   Methods & Data

To be able to tackle the given research question, the authors want to record the spread of a news story. They do that by constructing a dataset that consists of the main French news media for the whole of the year 2013. This includes newspapers, television channels, radio stations, pure online media and news agencies.

To be able to track a news story, one has to be able to detect a news story. To be able to achieve that, the authors develop a topic detection algorithm that identifies each news event. They basically do that by creating a class for each news topic and then just go back on the time axis.

The first story to break a news event is considered the original story. They then use their plagiarism algorithm to detect copies whereas they don't consider stories based on articles from news agencies copies, since the purpose of a news agency is to give a news event to media outlets.

To measure the impact of a news story, they track statistics like shares and likes on Facebook as well as Twitter.

4   Issue, Comments & Improvements

Impact Measure   First I want to point out, that it is hard to judge if their impact measure is actually worth while. The problem is, that they can only view public information yet the fact that it is public might be a bias in itself. It might be, that people have an agenda when sharing a news article

publicly. For example, if I read a good article, I rather share it directly i.e. privately with someone rather than posting it publicly but if I wanted to generate awareness for a specific topic, I'd post it publicly.

I also want to add, that there are studies showing that most of the articles that are being shared online aren't being read. The only thing that was read, was the title. Reinforcing my argument about an agenda based sharing desicion.

New Structrues Since this is a response essay and not an academic work, I'd like to think about some more controversial topics.

To me, it seems that the authors have a very strong bias towards the classical order of society resp. the structure of the fourth estate. The fourth estate has always been and will always be abused to lobby voters. So the current structure of article based news might simply not be a good model to distribute news. Working towards stronger copyright laws might harden this structure and in turn might keep us locked to the old way.

It might be, that we are better off by letting the news be as free as possible. Free meaning that there won't be a lot of copyright protection in place. This might allow for new technologies and structure to emerge like a personlized news assistant. A chat bot that derlivers the news to you. Such a tool could be checked for biases much easier than a normal news paper.

While it is good to think about how to preserve the status-quo, I also think it might be worthwhile to propose a very different solution. Maybe newspaper simply don't work in the digital age because the way young people consume news media might change drastically. Instead of reading an article, they probably much rather watch a YouTueb video about a selected topic or maybe even watch a TikTok.

To reiterate: Of course, the point of the paper wasn't to investigate alternative news distribution models but rather investigate the current one. But in a next step they do use their results to make a conclusion and that's where the bias takes place. I think that the conclusion could have been a good opportunity to take a broader stance.

Disregard of Youth Their impact measure relies on public statistics from Facebook and Twitter. This might disregard the youth because most young

people don't use Facebook and it's questionable how many use Twitter. Furthermore, their way of informing themselves might not be by reading an article, but by watching a video on YouTube or even TikTok.

While their dataset and impact measure might fit together, both seem to disregard the youth.