**Media in a Democracy**

The term “news” describes externally generated information about anything, but primarily about recent or current events. News delivers information about everything outside of a person’s personal sphere. Therefore, news plays a pivotal role in a democracy regarding the need to thoroughly inform citizens about the happenings of the entire nation to allow for intelligent election cycles and an optimally running government. The downfall of a democracy has consistently been an ignorant or easily manipulated general public. Historic philosophers have described democracy as flawed by the ignorance or disinterest of the masses. So considering the enormous stakes, how should news be optimally distributed to citizens of a democracy to allow for an educated and informed society?

As a preface, the problem statement would like to use algorithms to solve this problem, which makes some sense. However, the idea of an algorithm is very dangerous. Computer science certainly has a capability to sort and distribute news. Computer science has near-infinite capabilities, so any designed algorithm will generally be easily created. However, an improperly designed algorithm can have disastrous consequences, a la the housing crash in 2008 or racist financial (bail/loan) algorithms. This paper will primarily focus on building a nuanced logical philosophical foundation upon which we build our algorithm.

With regards to the scope of our project, the current complaints about the distribution of news is about political news. Thus, our scope is politics. However, everything is political. Considering the current impact of BLM on sports and the highly politicized nature of BLM, even sports are political. However, when I write political news, it just means news. We will use our model to sort news from the United Kingdom, but our logical basis will be built off of an analysis of American media and its relationship with American democracy.

How should political news be optimally distributed to citizens of a democracy to create an informed society? A full answer to this will take more time than is available. However, a basic optimal system of distribution is to expose citizens to a vast political spectrum of independently created and “correctly” motivated articles. The “correct” motivation for the majority of news should be to inform. It should be journalism for the sake of journalism. With respect to purely opinion articles, they are certainly required for a well-functioning democracy, and bias is admissible if the audience is exposed to a variety of biases. Audiences must somehow be open and accepting of all content provided, and have a trustworthy metric to determine the reliability and bias of a source.

How can we work from our current form of distribution towards the optimal one? Clearly, the status quo is very flawed. It is argued that people are currently trapped inside an echo chamber of sorts, where the news they are exposed to serves to reinforce pre-existing opinions. News is distributed to people who would already agree or support the content of the article. The stray articles that the person doesn’t agree with are discarded or ignored on the basis of the media outlet’s assumed bias. Additionally, the production of news itself is problematic. Major media outlets often have significant corporate or political interests, which causes its news to be similarly biased. With the rise of social media and the importance of strong online presences, media outlets are also incentivized to sensationalize their biases to increase their web traffic. These major media outlets then build one-dimensional audiences of one type or another, and must continue to pander to these audiences whose opinions they’ve shaped. As the audiences conform to the two-party political landscape of the US, political media itself becomes dichotomized into two streams of information. Entry into politics has been gated by funding for so long because candidates and parties with low funding aren’t broadcasted to the general public and are thus incapable of getting enough votes to win elections. What percentage of Americans are aware of the platform of Jo Jorgensen, the candidate from the third most popular party in the presidential election? Thus, we see three problems. People are trapped in an echo chamber, media is biased and sensationalized, and media is dichotomized.

So what’s the solution? Ideally our algorithm solves the echo chamber and dichotomy problem and exposes audiences to a large variety of journalism. The algorithm would provide a broad spectrum of political ideas regardless of popularity. Understandably, an appreciation of the spectrum would likely require an improved public education system, but that’s not within the scope of the project. Our algorithm to solve this is simple. We separate media outlets into groups based on bias, find the most reliable and accurate source(s) in each group, then give the audience news from these sources. Our algorithm scrapes the most recent politics headlines from 11 of the most popular UK news sites, then uses tensorflow and a clustering algorithm to sort newspapers by similar headlines. This model makes the assumption that similar newspapers cover similar events. We recognize that this algorithm has flaws in that the content of two articles with the same headline may have completely different biases, but an integral part of journalism is choosing what and what not to cover, and this aspect would be captured by our algorithm.

To solve the problem of corporate bias, political bias, and sensationalism, our algorithm will find a way to incentivize journalism for the sake of journalism. We cannot simply remove these types of journalism from society with the exception of extreme structural change, but we can reduce web traffic by preventing poor journalism from appearing in a search result. He who controls web traffic controls revenue streams and directly shifts the way journalism is produced. Media outlets, just like all corporations, will always pursue the interests of capital. If the interests of capital are to be reliable and accurate, then media outlets will be reliable and accurate. The algorithm will try to assign what are essentially accreditations to each source and only return the most reliable sources.

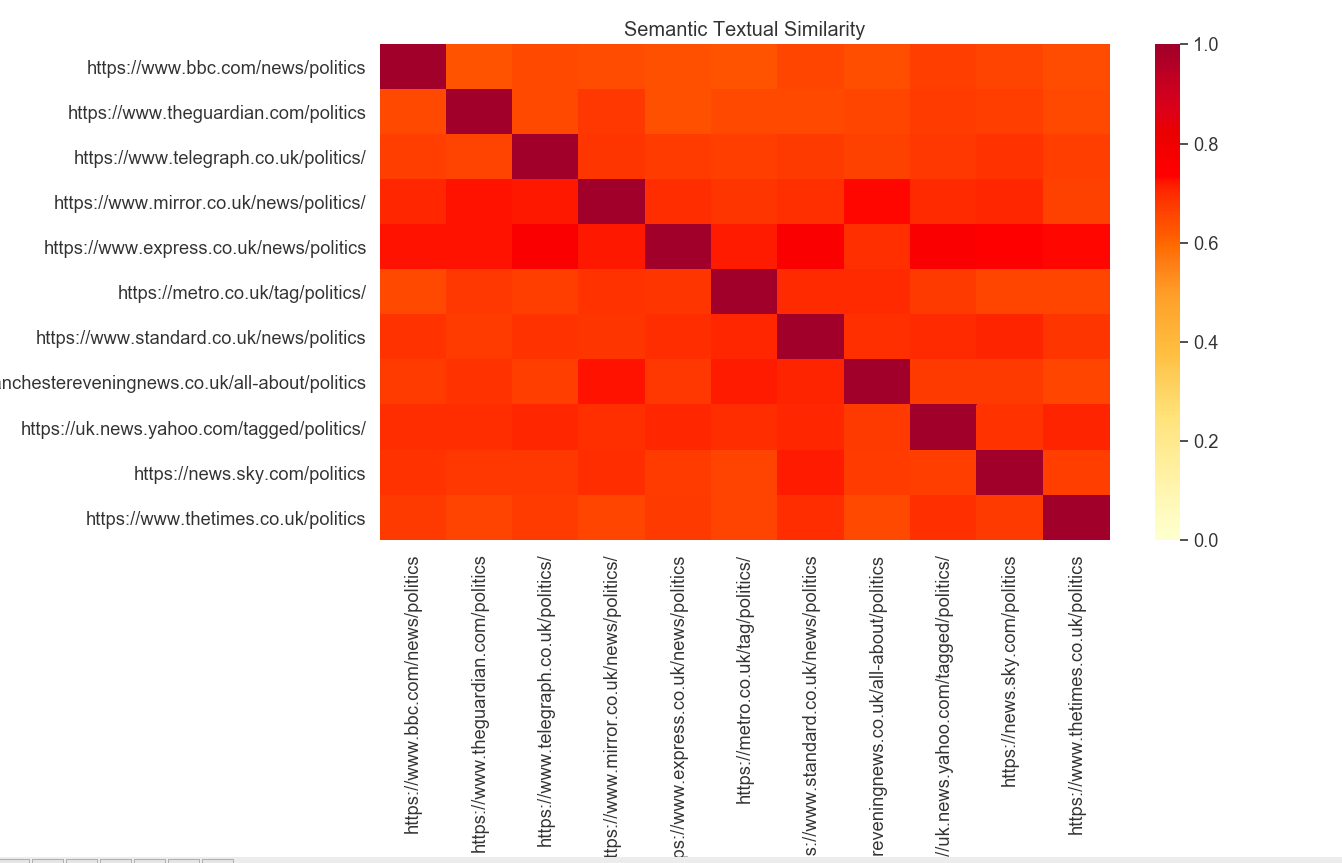
We considered using an algorithm to solve this accuracy problem, but algorithms seem too flawed. There is a remarkable amount of nuance to journalism that is impossible to capture through simple fact checkers. Tone, the ways facts are contextualized, and content writers choose to leave out are all journalism nuances that are difficult to analyze with the most advanced algorithm. For example, if we tried to analyze a website called domesticterrorism.com but the website only covered middle-eastern terrorist attacks, it wouldn’t necessarily be factually incorrect, but it would be incredibly flawed due to the content it intentionally chose to ignore. However, a fact checker would give it a perfect score, and would thus support an source whose popularization would have dangerous consequences

Although it’s arguable whether it’s possible for a fact-checker or accreditation third-party to be totally free of bias, there exist award panels such as the Pulitzer Prize which distribute awards for the best writing without much complaint. However, successful award panels such as the Pulitzer Prize are successful because they use humans to decipher the nuances and intricacies of language. Thus, although it isn’t an algorithm, we would ideally use an independent third-party organization of field professionals to give accreditations to different sources. There is only a finite number of media outlets and journalists, so it’s not an impossible task to use a human panel to cover them all.

Our “algorithm” was to choose the most reliable source in each cluster according to the independent third-party organization (in this case we decided to use mediafactbias.com) and return their news. Thus, we hope this algorithm which provides the most reliable sources across the political spectrum creates a system that prioritizes reliability and diversity.

The results from our current algorithm are slightly mixed due to a small sample size. We only covered the most popular UK newspapers (with the exception of daily mail because of their absurdly long headlines), and we only covered headlines that were on their politics pages which was a small sample size. This algorithm is quick and easily scalable to include more headlines and news sources. In the future, we can use headline archives and compare daily headline releases for more accuracy. However, the algorithm still managed to create a cluster of only left-leaning sources and a cluster of mostly right-leaning sources. Ideally, eventually, it wouldn’t be a cluster on a one-dimensional plane, but the political and thus media landscape is what it is.

Our results and some helpful graphics are below.



Above is the textual similarity of headlines on these news websites. Our clustering algorithm split these websites into two clusters. Our code and scraped headline data are in the attached repository

Cluster 1: Daily Express, Yahoo News, The Evening Standard, Telegraph, Sky News, The Times, BBC news

Cluster 2: Mirror, Manchester Evening News, the Guardian, Metro

Ranked clusters are in the spreadsheet.