9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Mapping the U.S. News Landscape

Keywords: online news, media consolidation, local newspapers

Extended Abstract

Over the last decade, local newspapers in the U.S. have seen a decline in revenue and readership. With the rise of digital information sources, print newspapers have found themselves losing advertising and subscription revenue, which has in turn led to a decline in newsroom employment [1] and, in many cases, the eradication of the newspapers altogether, with those surviving being bought up by large media companies that run them with a skeletal staff and reduced printing [2]. Simultaneously, significant attention has been placed on the phenomenon of new digital sites spreading misinformation while presenting themselves as reputable newspapers. Known as pink-slime journalism, these websites are run by media companies seeking to publish politically biased news stories under the guise of local news. While there exist multiple databases of newspapers and news sites in the U.S., each of these databases captures a different aspect of the news ecosystem. This raises questions such as: where in the U.S. have local newspapers been disappearing, and where have digital sites moved in to capture their audience? How consolidated is the media landscape, and does the owner have an impact on the political stance of the publication? What counties and cities have the fewest newspapers per capita, and how does this relate with their voting history? And how much agreement is there between the bias and credibility labels provided in different databases?

In this paper, we analyze the state of the U.S. news landscape by collecting and integrating newspaper metadata from multiple sources into one comprehensive database which contains information on the news sites' location, owner, bias, credibility, consumption, and numerous other features. For this purpose of this analysis, we currently only focus on news websites that primarily publish written content, and not TV channels and radio stations (which may also have websites of their own). Table 1 lists the datasets used in our work.

Dataset	# Domains
News Deserts[1]	4863
NewsGuard[2]	5795
Media Bias / Fact Check (MBFC)[3]	3642
AllSides[4]	456
Paperboy[5]	4482
Union	**12731**
Intersection	**46**

Table 1: Statistics on the datasets and their intersection and union.

[1] usnewsdeserts.com
[2] newsguardtech.com
[3] mediabiasfactcheck.com
[4] allsides.com
[5] thepaperboy.com

Figure 1 shows a summary of the commonalities and differences between the five datasets. We see from subfigure 1a that each dataset captures only a certain subset of attributes, with two types of datasets becoming apparent. In the first type are News Deserts and Paperboy, which document the names and geographical location of local and regional newspapers in the U.S. Paperboy additionally gives us their social media links, while News Deserts instead gives us information on their owner and circulation numbers. In the other type of dataset, we have NewsGuard, AllSides, and MBFC, which are focused primarily on rating the political bias and journalistic credibility of news sources, and thus contain the relevant fields for that. Clearly, no dataset captures all of the information necessary for charting the news landscape properly.

Furthermore, even when we consider only datasets of similar nature, we find a low overlap in terms of the newspapers they contain. In order to maximize matches, we clean all website URLs and keep only the core domain names. Even though News Deserts and Paperboy both focus on local newspapers, they share a little over half of them in common. The same is true for NewsGuard and MBFC, both of which focus on gauging news websites' slant and quality. Clearly, no single dataset is complete on its own, and we can only paint a full picture of the U.S. news landscape, both in terms of coverage (news websites) and detail (metadata), by combining all of them together. Table 1 shows the statistics of the datasets and their intersection and union.

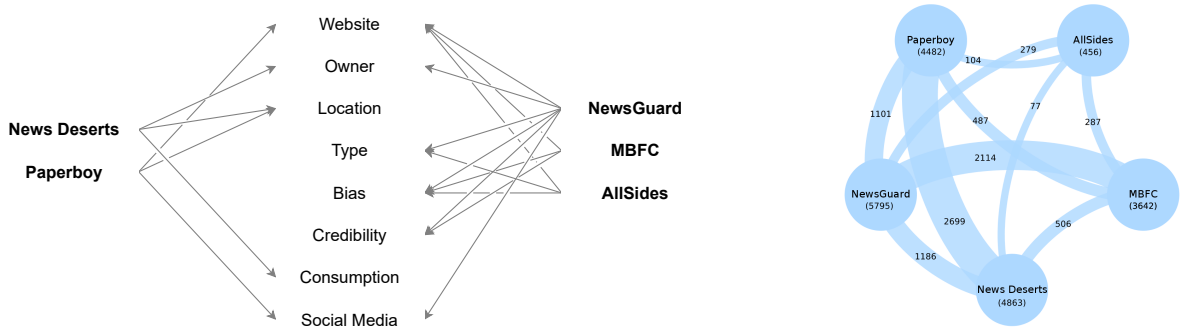
To further understand the state of news, Figure 2a shows a view of the media consolidation that has been on the rise over the last decade. One-third of all newspapers (and two-thirds of all daily newspapers) are owned by the 30 largest media companies. Even among these 30 companies, the distribution is very uneven: Gannet/GateHouse, a conglomerate formed by the merger of the two media companies in 2019, owns almost 1 in 10 of all newspapers in the U.S. (1 in 5 dailies), raising concerns about its impact on local newspapers and journalism. Figure 3 shows the geographic distribution of these newspapers. Using the News Deserts and Paperboy datasets, we map the number of newspapers in each state per million residents. We see that states with more newspapers are naturally those with higher populations as well, while those with lower populations have more newspapers per capita. In Figure 4, we compare the websites' bias ratings and credibility scores. Subfigure 4a shows bias scores assigned to 289 news publishers listed in all three datasets with bias ratings: NewsGuard, AllSides, and Media Bias / Fact Check (MBFC). We see a similar distribution across all three, with more websites marked as left-leaning than right-leaning, and AllSides and MBFC having a higher correlation with each other than the other two pairings. Subfigure 4b shows the correlation between bias and trustworthiness according to NewsGuard. We see that websites on the right are less likely to be trustworthy than those on the right or on the center.

We aim to extend this work to study the interplay between more factors like ownership, geographic location, political bias, and print circulation and online consumption. We hope that this work encourages further research on the news media in a comprehensive manner by looking at multiple sources of news websites and metadata on them in order to paint a more complete picture of the news landscape.

References

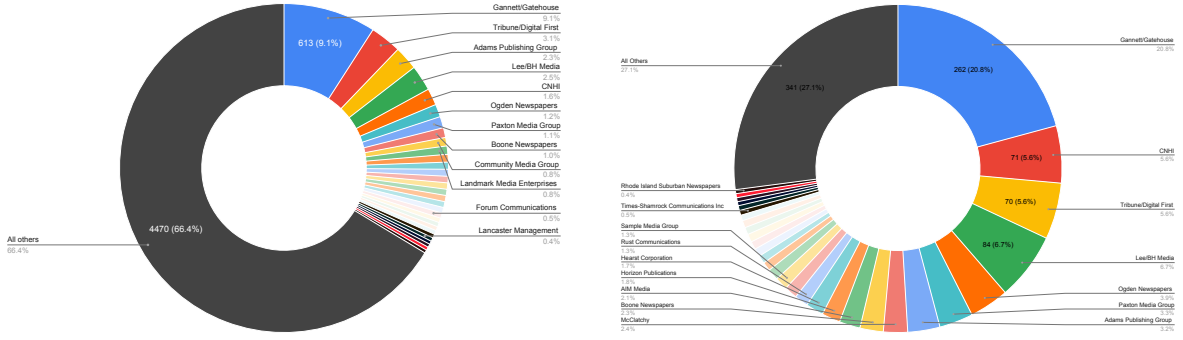
[1] E. Grieco, "U.S. Newspapers Have Shed Half of Their Newsroom Employees Since 2008," *Pew Research*, 2020.

[2] P. M. Abernathy, *News Deserts and Ghost Newspapers: Will Local News Survive?* Center for Innovation and Sustainability in Local Media, School of Media and Journalism, University of North Carolina at Chapel Hill, 2020.



(a) (b)

Figure 1: A comparison of all five databases. Subfigure (a) shows a condensed schema of the datasets used in our comparison. The middle column lists the common fields found in the different datasets, and arrows leading to them from the datasets listed on the sides show which fields each contains. We see that the the datasets on the left (News Deserts, Paperboy), are more similar as they contain information on local and regional newspapers. On the other hand, the datasets on the right (NewsGuard, MBFC, AllSides) are all focused on categorizing news sources based on their bias and type of publication. Subfigure (b) shows a network depicting the overlap of news websites contained in each of the five databases. We see that databases which aim to capture similar aspect of the news landscape share more in common with each other: Paperboy and News Deserts document traditional local and regional newspapers, while NewsGuard and MBFC document mostly online news publishers that may not print editions.



(a) (b)

Figure 2: A view of the state of media consolidation showing the total number of newspapers owned by each of the major media companies in the U.S. (a), and the number of dailies owned by them (b). One-third of all newspapers (and two-thirds of all dailies) are owned by the top 30 media companies. We see that Gannet/GateHouse is the largest by far and owns over 600 newspapers (over 250 dailies).

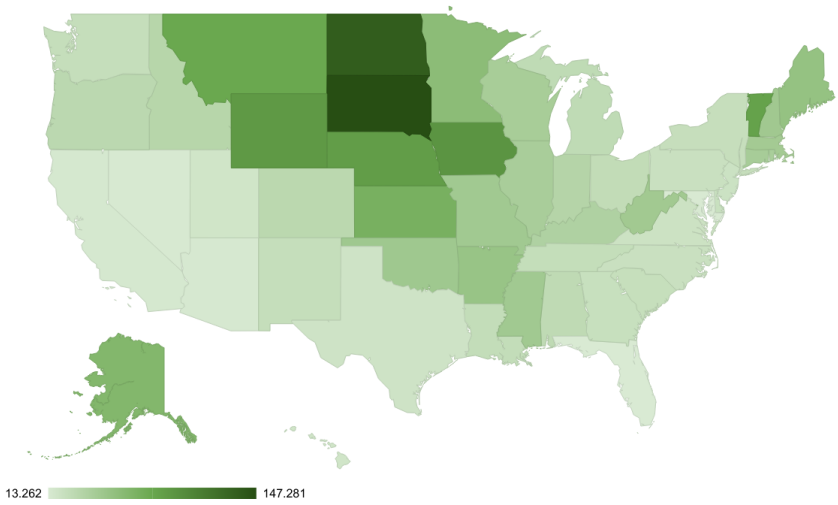
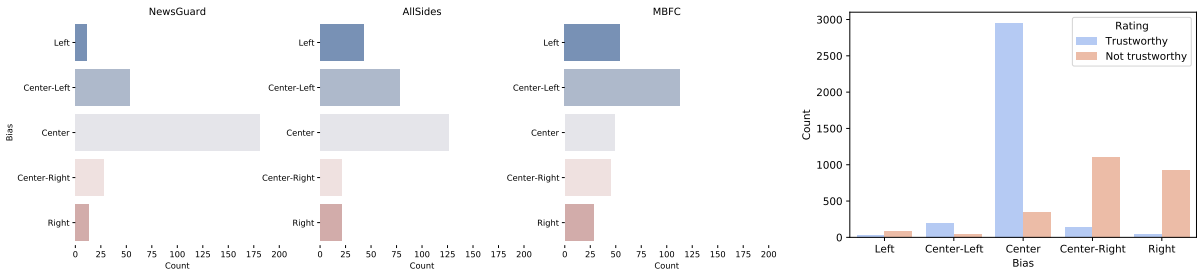


Figure 3: A map of the distribution of local and regional newspapers per million people for each state, using combined data from News Deserts and Paperboy. Florida has the lowest number of newspapers by population at 13 per 1M people, while South Dakota has the highest at 147 per 1M people.



(a) (b)

Figure 4: A comparison of political bias scores and the correlation between bias and credibility. Subfigure (a) shows bias scores assigned to the same set of 289 publishers by three different rating sources: NewsGuard, AllSides, and Media Bias / Fact Check (MBFC). NewsGuard and AllSides share the most similarity with each other: when mapped to a linear numerical scale in the range [-1, 1], the correlation of AllSides with MBFC is 0.5, while that of AllSides and MBFC with NewsGuard is 0.14 and 0.13, respectively. Subfigure (b) shows the relation between bias and trustworthiness scores according to NewsGuard. We see that websites publishing right-leaning content are less likely to be trustworthy.