Political audience diversity and news quality

Shun Yamaya
Dartmouth College

Saumya Bhadani
University of South Florida

Alessandro Flammini
Indiana University Bloomington

Filippo Menczer
Indiana University Bloomington

Brendan Nyhan
Dartmouth College

Giovanni Luca Ciampaglia*
University of South Florida

ABSTRACT

A key factor in the dissemination of inauthentic information on social media is the interaction between human cognitive bias and algorithmic content curation. Research shows that some people have strong preferences for pro-attitudinal content. These tendencies are exacerbated by social media platforms, which often amplify low-quality content because it generates high levels of engagement among like-minded audiences. We propose using audience diversity as a quality signal to address this problem. To demonstrate the potential value of this approach, we combine a comprehensive dataset of news source reliability ratings compiled by domain experts with the web browsing histories from samples of U.S.-based Internet users. Our results indicate that partisan audience diversity is a potentially useful signal of higher journalistic standards. These results suggest that platforms should consider incorporating audience diversity into algorithmic ranking decisions.

KEYWORDS

Audience partisan diversity, news quality, content recommendation, YouGov Pulse Panel, NewsGuard.

1 INTRODUCTION

Search and recommendation algorithms shape the information that people see online. Conversely, the online activity of both producers and consumers of political news affects the design and evolution of algorithmic systems such as social media platforms. Unfortunately, not enough is known about how these processes interact or how this interaction affects the quality of the content that is presented to the public.

A particular concern is the recent, unforeseen explosion of inaccurate and inauthentic political information on the feeds of the major social media platforms [11], which may be the result of the interaction between human and algorithmic behavior. People tend to prefer pro-attitudinal information and to choose it when given the option, in a process that is called selective exposure [9]. Americans' information diets are less affected by this tendency in practice than many assume [6], but the people who consume the most political news are those mostly affected by this behavior. As a result, the news audience is far more polarized than the public as a whole [7]. Low-quality or false news that appeal to these tendencies may thus generate high levels of readership or engagement among narrow audiences online, prompting algorithms that seek to maximize engagement to distribute that content more widely.

Prior research indicates that recommendation algorithms may indeed show a tendency to promote items that have already achieved popularity [14]. This form of "popularity" bias can in turn influence the overall quality of information consumed by users [3, 10], perhaps even in counter-intuitive ways [4]. In particular, news recommendation systems affected by popularity bias may be particularly vulnerable to automated amplifiers, which could exploit the inclination to spread low quality content to like-minded audiences [16].

To counter these tendencies, social media platforms are seeking to include signals about the quality of news producers in their content recommendation algorithms. There is a vast literature about assessing the credibility of online content [2, 8] or the reputation of individual online users [1, 5]. Unfortunately, many of these methods are hard to scale and/or highly sensitive to context or to the type of content being generated (e.g., wikis). As a result, they may not easily generalize to the news domain. One approach is to try to evaluate the quality of websites directly [19], but such an approach is costly to scale and may fail to keep up with new sources of information that appear on platforms. Alternatively, platforms could rely on crowdsourcing. While research shows that news consumers are generally able to distinguish between high and low quality news sources [15], crowdsourced signals are also vulnerable to manipulation as well as delays in evaluating new sources.

In this work, we propose using the partisan diversity of a website's audience as a quality signal. This approach has two key advantages. First, it is easy to compute at scale given that information about the partisanship of some users is known. In addition, it is less prone to manipulation by automated amplifiers if one can detect inauthentic accounts [17, 18]. Both conditions are easily satisfied by social media platforms given the wealth of user information they routinely collect.

We evaluate our proposed approach by combining two data sources: a comprehensive data set of web traffic history from 19,298 American citizens (collected as part of surveys of the YouGov Pulse panel), and 3,765 credibility ratings of web

*Corresponding author. Email: glc3@usf.edu

domains by expert journalists (compiled by NewsGuard). As a preview of our results, we first establish that the number of pageviews that a domain receives is not associated with the overall journalistic quality of the website — in other words, popularity does not predict quality, which highlights the potential problem with algorithmic recommendation systems. Next, we introduce a variety of measures that operationalize the concept of partisan diversity of a website's audience and show that these measures predict website quality better than popularity. Our results suggests that partisan audience diversity metrics could be a useful signal to improve the quality of news sources on social media.

2 METHODS

To study the online behavior of humans and the quality of the websites they visit, we bring together two sources of data. The first is the NewsGuard News Website Reliability Index, a list of web domain reliability ratings compiled by a team of professional journalists and news editors. To date, NewsGuard has rated 3,765 web domains on a 100-point scale based on a number of journalistic criteria such as editorial responsibility, accountability, and financial transparency.[1] NewsGuard categorizes web domains into four main groups: "Green" domains, which have a score of 60 or more points, are considered reliable; "Red" domains, which score less than 60 points, are considered unreliable; "Satire" domains, which should not be regarded as news sources regardless of the score; and "Platform" domains, like Facebook or YouTube, which primarily host content generated by their users rather than producing their own news. The mean reliability score is 69.6; the distribution of scores is shown in Fig. 1.

The second data source is the YouGov Pulse panel, a sample of U.S.-based Internet users whose web traffic was collected in anonymized form with their prior consent. This traffic data was collected during seven periods between October 2016 and March 2019 (see Table 1). A total of 19,298 participants provided data. In addition to their web traffic logs, participants reported their partisanship on a seven-point scale in online surveys. We pool web traffic for each domain that received 30 or more unique visitors and use the self-reported partisanship of the visitors to estimate average visitor partisanship, which we refer to as partisan slant, and audience diversity, which we estimate using different measures described further below.

To estimate audience diversity, let us consider a generic web domain, and let's define as N_k the count of participants who visited that domain and reported their political affiliation to be equal to k for $k = 1, \ldots, 7$ (where 1 = strong Democrat and 7 = strong Republican). The total number of participants who visited the domain is thus $N = \sum_k N_k$, and the fraction of participants with partisanship value k is $p_k = N_k/N$. Let us also denote with s_i the partisanship of the i-th individual. We can define audience diversity in different ways. Here we use the following metrics:

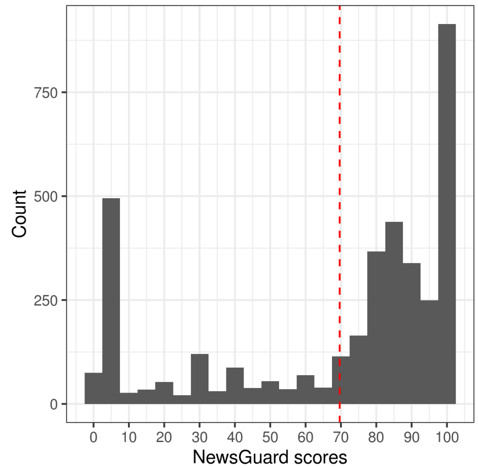

Figure 1: Distribution of NewsGuard scores (N = 3,765). The red dashed line indicates the average score.

(1) *Variance*: $\sigma^2 = N^{-1} \sum (s_i - \bar{s})^2$, where $\bar{s}$ is average partisanship;
(2) *Shannon's Entropy*: $S = -\sum p(k) \log p(k)$, which we estimate in several ways (see below);
(3) *Inverse Max-Prob*: $1 - \max_k \{p_k\}$;
(4) *Inverse Gini*: $1 - G$ where G is the Gini coefficient of the count distribution $\{N_k\}_{k=1\ldots7}$.

Shannon's Entropy requires an estimate of $p(k)$, the probability that a visitor has partisanship k. We use the following approaches to estimate this quantity:

(1) *Maximum Likelihood*, or $p(k) = p_k$;
(2) *Dirichlet*, or $p(k) = \frac{N_k+\alpha}{N+7\alpha}$. This corresponds to the mean of the posterior probability computed with a Dirichlet prior with $\alpha = 1$;
(3) *Nemenman, Shafee and Bialek (NSB)* approach [13]. In this approach, a mixture of different Dirichlet priors is employed to get a relatively unbiased prior for calculating the posterior probability of each partisan slant.

The above metrics all capture the intuitive idea that the political diversity of the audience of a web domain will be reflected by its distribution of traffic across different partisan groups. They do so by considering the contribution of each individual visitor evenly and thus can be regarded as user-level measures of diversity. However, the volume and content of web browsing activity is highly heterogeneous across internet users [7, 12]. We therefore also compute equivalent measures of audience diversity at the level of individual

[1]These data were current as of November 12, 2019 and do not reflect subsequent updates made after that time.

Table 1: YouGov Pulse respondent data summary.

Duration	Respondents	Domains	Pageviews
Oct. 07, 2016 – Nov. 14, 2016	3,251	158,706	26,715,631
Oct. 25, 2017 – Nov. 21, 2017	2,100	104,513	14,247,987
Jun. 11, 2018 – Jul. 31, 2018	1,718	108,953	15,212,281
Jul. 12, 2018 – Aug. 02, 2018	2,000	74,469	9,395,659
Oct. 05, 2018 – Nov. 05, 2018	3,332	98,850	19,288,382
Nov. 12, 2018 – Jan. 16, 2019	4,907	117,510	21,093,638
Jan. 24, 2019 – Mar. 11, 2019	2,000	113,700	27,482,462

pageviews rather than the user-level metrics described above. These measures weight repeat visitors to a web domain in proportion to how frequently they visited it rather than assigning the same weight to each individual.

3 RESULTS

To motivate our study, we first demonstrate the problem with algorithms prioritizing content that attracts large audiences. To do so, we examine the relationship between audience size and website quality in the YouGov Pulse data. Due to skew in the audience size among domains, we analyze audience size on a logarithmic scale. Fig. 2 shows that the amount of traffic that a website attracts is not associated to the quality of its journalistic practices, which we measure using NewsGuard scores. The lack of association persists even if we separately consider websites with predominantly Democratic or Republican audiences (details available upon request).

In contrast, we observe that sites with high levels of audience diversity tend to score higher on the NewsGuard quality metric, whereas those with highly partisan audiences and correspondingly low levels of diversity tend to score lower. Fig. 3 provides a visualization of this relationship using average audience partisanship and partisan audience diversity at the website level. In this figure and those that follow, the diversity metric employed is the variance. It is immediately noticeable that many of the lowest quality sites are in the tails of the distribution due to having highly slanted audiences with low levels of audience diversity. This effect is especially pronounced on the right side of the figure, which corresponds to the sites with largely Republican audiences.[2]

We plot the diversity–quality relationship more formally in Fig. 4, which shows that audience partisan diversity is positively associated with news quality. This relationship holds both at the user level (left panel) and at the pageview level (right panel), but is stronger at the user level (i.e., when our results are not weighted by the number of times a website is visited by a given participant). The relationship is also stronger for sites whose average visitor identifies as a Republican versus a Democrat.

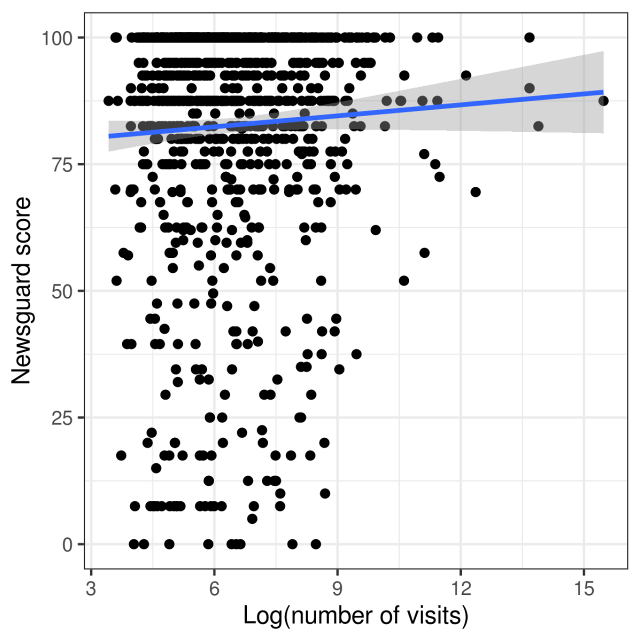

Figure 2: Relationship between audience size and news quality by domain. The Pearson correlation is 0.05.

Finally, we consider the problem of choosing the correct operational definition of audience diversity. We repeat the above analysis for all diversity metrics presented above and summarize the results in Table 2. For each metric, we estimate the degree of linear association with news quality using the Pearson correlation coefficient. We also report the R^2 coefficient of determination and the p-value of the F-statistic as a measure of significance of the fit. Each metric is positively correlated with quality at the user level, but we find that the relationship is strongest for variance of audience

[2]Fig. 3 also shows that the partisan diversity of an audience relates to its average partisanship in an inverse U-shaped pattern. This could suggest that average audience partisanship is associated with news quality, but could also just be a mere consequence of our use of an ordinal scale to measure partisanship. Our focus here is audience diversity, but future work should investigate this potential association.

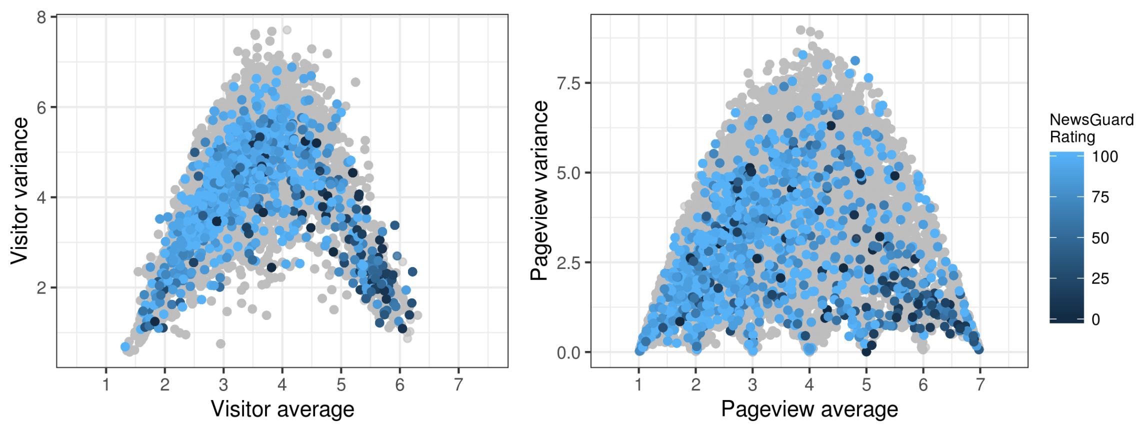

Figure 3: Average audience partisanship versus variance. Left panel: individual users. Right panel: weighted by pageviews. Domains with a news quality score are shaded in blue (where darker shades equal lower scores). Domains with no score are plotted in gray.

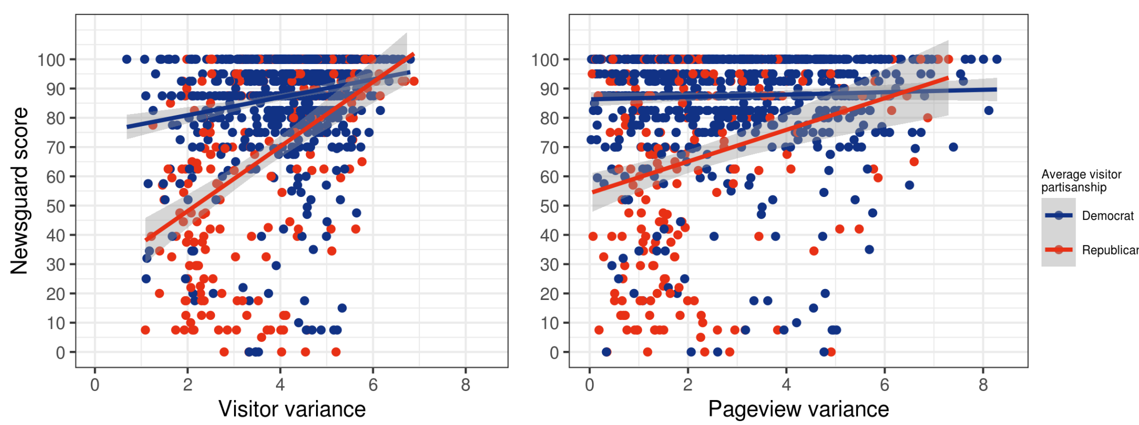

Figure 4: Relationship between audience partisan diversity and news quality for websites whose average visitor is a Democrat or a Republican. Left panel: individual users. Right panel: weighted by pageviews.

partisanship. At the pageview level, however, the association disappears for all metrics but variance, which still produces a modest correlation.

4 DISCUSSION

These results demonstrate that diversity in audience partisanship can serve as a useful signal of news quality at the domain level. Using data on millions of pageviews from U.S.-based Internet users, we show that the variance of partisanship among all unique visitors to a website is strongly positively associated with the scores for news quality created by NewsGuard. These relationships are weaker for other measures of diversity or if we weight respondent partisanship by pageviews in calculating audience partisan diversity. Finally, we observe that the diversity-quality relationship is stronger among sites whose audiences lean Republican compared to those whose audiences lean Democratic.

Future research should investigate the reasons why audience diversity predicts news quality more poorly when weighted by pageviews and why the strength of the relationship differs between outlets with Democratic- and Republican-leaning audiences. It is also important to evaluate how strongly audience diversity predicts news quality when accounting for total audience size, to conduct out-of-sample tests of the predictive power of this approach, and to go beyond the domain level to examine specific subdomains and pages that are focused on hard news topics.

Nonetheless, these results provide promising new evidence that audience partisan diversity can help platforms identify quality websites. We also hope that this metric could inform the evaluation of new sites and help algorithms avoid

Table 2: Relationship between audience partisan diversity and news quality.

Diversity metric	Correlation	R^2	p-value
USER LEVEL			
Variance	0.32	0.10	< 0.01
Entropy (Dir.)	0.21	0.04	< 0.01
Entropy (ML)	0.20	0.04	< 0.01
Entropy (NSB)	0.22	0.05	< 0.01
Inv. MaxProb	0.06	0.00	0.03
Inv. Gini	0.14	0.02	< 0.01
PAGEVIEW LEVEL			
Variance	0.14	0.02	< 0.01
Entropy (Dir.)	-0.02	0.00	0.58
Entropy (ML)	-0.02	0.00	0.62
Entropy (NSB)	-0.02	0.00	0.61
Inv. MaxProb	-0.04	0.00	0.14
Inv. Gini	-0.03	0.00	0.26

recommending those who attract highly skewed audiences. In this way, our approach could help to diminish the incentives that encourage entrepreneurs to create untrustworthy partisan websites in the first place.

Acknowledgements

We thanks NewsGuard for licensing the data and acknowledge Andrew Guess and Jason Reifler, Nyhan's coauthors on the research project that generated the web traffic data used in this study. This work was supported in part by the National Science Foundation under Grant No. 1915833.

REFERENCES

[1] B. Thomas Adler and Luca de Alfaro. 2007. A Content-driven Reputation System for the Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 261–270. https://doi.org/10.1145/1242572.1242608

[2] Jin-Hee Cho, Kevin Chan, and Sibel Adali. 2015. A Survey on Trust Modeling. *ACM Comput. Surv.* 48, 2, Article 28 (Oct. 2015), 40 pages. https://doi.org/10.1145/2815595

[3] Giovanni Luca Ciampaglia, Azadeh Nematzadeh, Filippo Menczer, and Alessandro Flammini. 2018. How algorithmic popularity bias hinders or promotes quality. *Scientific Reports* 8, 1 (2018), 15951–. https://doi.org/10.1038/s41598-018-34203-2

[4] Fabrizio Germano, Vicenç Gómez, and Gaël Le Mens. 2019. The Few-get-richer: A Surprising Consequence of Popularity-based Rankings?. In *The World Wide Web Conference (WWW '19)*. ACM, New York, NY, USA, 2764–2770. https://doi.org/10.1145/3308558.3313693

[5] Jennifer Ann Golbeck. 2005. *Computing and applying trust in web-based social networks.* Ph.D. Dissertation. University of Maryland at College Park. http://hdl.handle.net/1903/2384

[6] Andrew Guess, Benjamin Lyons, Brendan Nyhan, and Jason Reifler. 2018. Avoiding the Echo Chamber about Echo Chambers: Why Selective Exposure to Like-minded Political News is Less Prevalent than you Think. (2018). Knight Foundation report, February 12, 2018. Downloaded October 23, 2018 from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf.

[7] Andrew M. Guess. 2018. (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. (2018). Unpublished manuscript. Downloaded September 10, 2018 from https://webspace.princeton.edu/users/aguess/Guess_OnlineMediaDiets.pdf.

[8] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. *TweetCred: Real-Time Credibility Assessment of Content on Twitter.* Springer International Publishing, Cham, 228–243. https://doi.org/10.1007/978-3-319-13734-6_16

[9] William Hart, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J. Lindberg, and Lisa Merrill. 2009. Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin* 135, 4 (2009), 555.

[10] Tadd Hogg and Kristina Lerman. 2015. Disentangling the Effects of Social Signals. *Human Computation* 2, 2 (2015), 189–208.

[11] David Lazer, Matthew Baum, Yochai Benkler, Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein, Emily Thorson, Duncan Watts, and Jonathan Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096. https://doi.org/10.1126/science.aao2998

[12] A. L. Montgomery and C. Faloutsos. 2001. Identifying Web browsing trends and patterns. *Computer* 34, 7 (July 2001), 94–95. https://doi.org/10.1109/2.933515

[13] Ilya Nemenman, Fariel Shafee, and William Bialek. 2001. Entropy and Inference, Revisited. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic (NIPS'01)*. MIT Press, Cambridge, MA, USA, 471–478. http://dl.acm.org/citation.cfm?id=2980539.2980601

[14] Dimitar Nikolov, Mounia Lalmas, Alessandro Flammini, and Filippo Menczer. 2019. Quantifying Biases in Online Information Exposure. *Journal of the Association for Information Science and Technology* 70, 3 (2019), 218–229. https://doi.org/10.1002/asi.24121

[15] Gordon Pennycook and David G. Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (2019), 2521–2526. https://doi.org/10.1073/pnas.1806781116

[16] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9, 1 (Nov. 2018), 4787. https://doi.org/10.1038/s41467-018-06930-7

[17] Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *Proc. Eleventh Intl AAAI Conference on Web and Social Media (ICWSM '17)*. AAAI, Palo Alto, Calif., USA, 280–289. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587

[18] Kai-Cheng Yang, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies* 1, 1 (2019), 48–61. https://doi.org/10.1002/hbe2.115

[19] Amy X. Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, Ed Bice, Sandro Hawke, David Karger, and An Xiao Mina. 2018. A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles. In *Companion Proc. of The Web Conference 2018 (WWW '18)*. Intl. World Wide Web Conf. Steering Committee, Republic and Canton of Geneva, Switzerland, 603–612. https://doi.org/10.1145/3184558.3188731