Cracking open the European Newsfeed over time

Keywords: problematic information, facebook, news consumption, fake news

Extended Abstract

This paper contributes to the ongoing effort to describe and quantify the quality of information that is shared on large social media platforms. We do this by complementing existing research that provided a first quantitative assessment of the quality of the information circulating on Facebook among US users. Leveraging an updated version of Meta's URL Shares dataset we quantify the trustworthy and untrustworthy links to external websites that have been shared on Facebook in the period between 2018 and 2021 in the US and in three major European countries (Germany, France and Italy). We observe how the amount of untrustworthy sources is - excluding few "exceptional" years - practically stable over time and comparable between European countries and the US. Additionally, we focus on how exposure and sharing of trustworthy and untrustworthy sources varies between different age-groups confirming the finding previously observed in the US that shows how exposure and engagement with untrustworthy sources grows with the age of the group we are considering.

Data and Methods:
To measure the exposure to news stories from trustworthy and untrustworthy news sources on Facebook, we relied on the URL Shares Dataset (Kifer et al., 2020) provided by Meta. Unlike other Facebook data sources, this dataset provides a set of user-centred metrics (including exposure) concerning the most popular URLs shared on the platform. The demographics of users who viewed and otherwise interacted with a URL are broken down by gender, age classes, and - only for US users - user's estimated political leaning. For our analysis we leverage the latest version that covers a period of time that goes from June 2017 to June 2022. Data is aggregated by month. Repeated exposure or interactions on the same URL by the same user are only counted ones. Given this peculiarity, the metrics presented are secured by differential privacy (D'Orazio et al. 2015), a privacy-preserving technique that adds noise to the data to avoid user re-identification. For our studies, we have categorized URLs into two broad groups based on the trustworthiness of the domains they belong to for each country/year under investigation. To assess the trustworthiness of domains, we utilized the findings of NewsGuard a private company that assigns a score ranging from 0 to 100 to each news domain they have evaluated, reflecting their adherence to a set of journalistic standards. Each domain is investigated by a team of specialized journalists and the scores are periodically reviewed. Total scores over 60 indicate a trustworthy news source (T). News sourcing scoring lower than 60 are instead labelled as untrustworthy (N). Filtering the data present in the URL Shares Datasets for the domains we have rated by Newsguard leaves us with 9526037 unique URLs from 1,314 domains having being shared in the 4 countries from 2018 to 2021.

Results
First we present the overall percentage of URLs from trustworthy and untrustworthy domains in the four countries (Figure 1).On a general level the numbers we observe are somehow comparable with the 18% observed for the 2018 US data by Guess and colleagues (2021).

When looking into the details, it is possible to observe general stability besides a few fluctuations. Fluctuations are mostly noticeable in Germany, where the percentage of links from untrustworthy domains reached the maximum in 2018 and 2021, with more than 21%. France shows a somewhat stable decline from 14.5% in 2018 to 12.7% in 2021 while Italy shows the opposite trend growing from 12.5 % in 2018 to 17.1% in 2021.
Second we measure exposure to trustworthy and untrustworthy links (Figure 2). We do that by quantifying the overall number of *views* obtained by the different types of URLs. Views of links from untrustworthy sources have been stable at approximately 5% in Germany over the entire period. Views of untrustworthy URLs show a declining trend in France and the US, with France showing a remarkable reduction from more than 10% in 2018 to 4.3% in 2021. Italy shows a different trend, with a peak of Views for untrustworthy content (8.4 %) reached in 2019.
Third we measure the sharing activity generated by trustworthy and untrustworthy links (Figure 3). The percentage of shares for links from untrustworthy sources over the years (Figure 3) has grown in Italy – from 9.6% in 2018 to 17.2 % in 2021–. In France, it shows a clear decline, from almost 13.4% in 2018 to 9.3% in 2021. In Germany, the shares percentage declined from 2018 to 2020 but reached the highest observed proportion in 2021 (22.6%). It should be noted that the actual number of URLs shared in the data declines dramatically from 2020 to 2021 in all the 4 countries. Looking at the U.S., the data is substantially stable between 13% and 15%.

Finally we look at the number of views and shares divided per the age group of the users. This allows us to understand if, as it has been claimed (Guess et al. 2021), the problem of unreliable information is more relevant for specific demographics. Figures 4 and 5 break down the number percentage of views and share for the different age groups. We can observe how the proportion of views of content from untrustworthy sources grows with the age of the group we take into consideration. While the trend is observable in all the countries, it appears remarkably stable in the US, where more than 14% of the news URLs shown to users older than 65 years old are from untrustworthy sources.
Looking at the proportion of trustworthy/untrustworthy links shared by the various age groups we see how the trend we observed about exposure is largely confirmed. The only noticeable exception appears to be in Germany in 2021, where the largest proportion of shares of untrustworthy content was observed among the content shared by 45-56 years old. Overall these data confirm that both online exposure and sharing or unreliable content seems to be connected with the age of the users we are considering.

References

D'Orazio, V., Honaker, J., & King, G. (2015). Differential privacy for social science inference. Sloan Foundation Economics Research Paper, (2676160).

Kifer, D., Messing, S., Roth, A., Thakurta, A., & Zhang, D. (2020). Guidelines for implementing and auditing differentially private systems. arXiv preprint arXiv:2002.04049.

Guess, A., Aslett, K., Tucker, J., Bonneau, R., & Nagler, J. (2021). Cracking open the news feed: Exploring what us Facebook users see and share with large-scale platform data. Journal of Quantitative Description: Digital Media, 1.

Figure 1: Percentage of links from trustworthy/untrustworthy sources shared in the 4 countries.

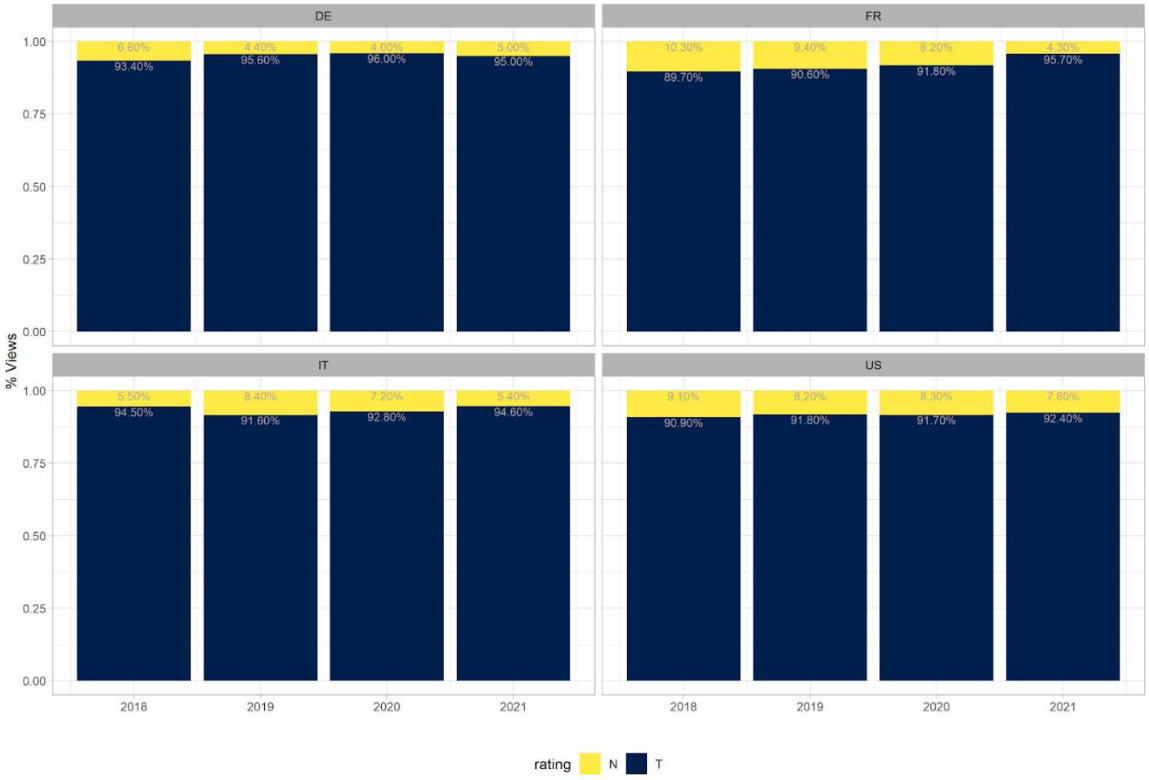

Figure 2 Percentage of Views received by trustworthy and untrustworthy links

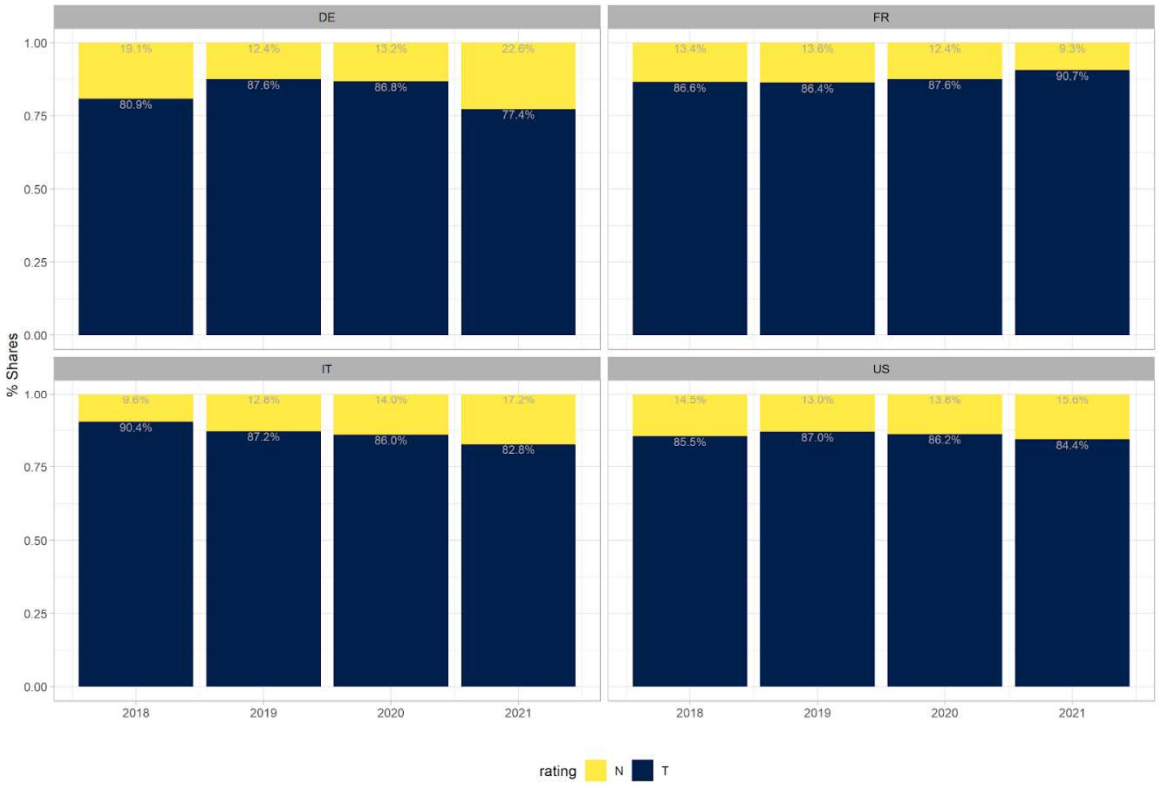

Figure 3: Percentage of Shares received by trustworthy and untrustworthy links

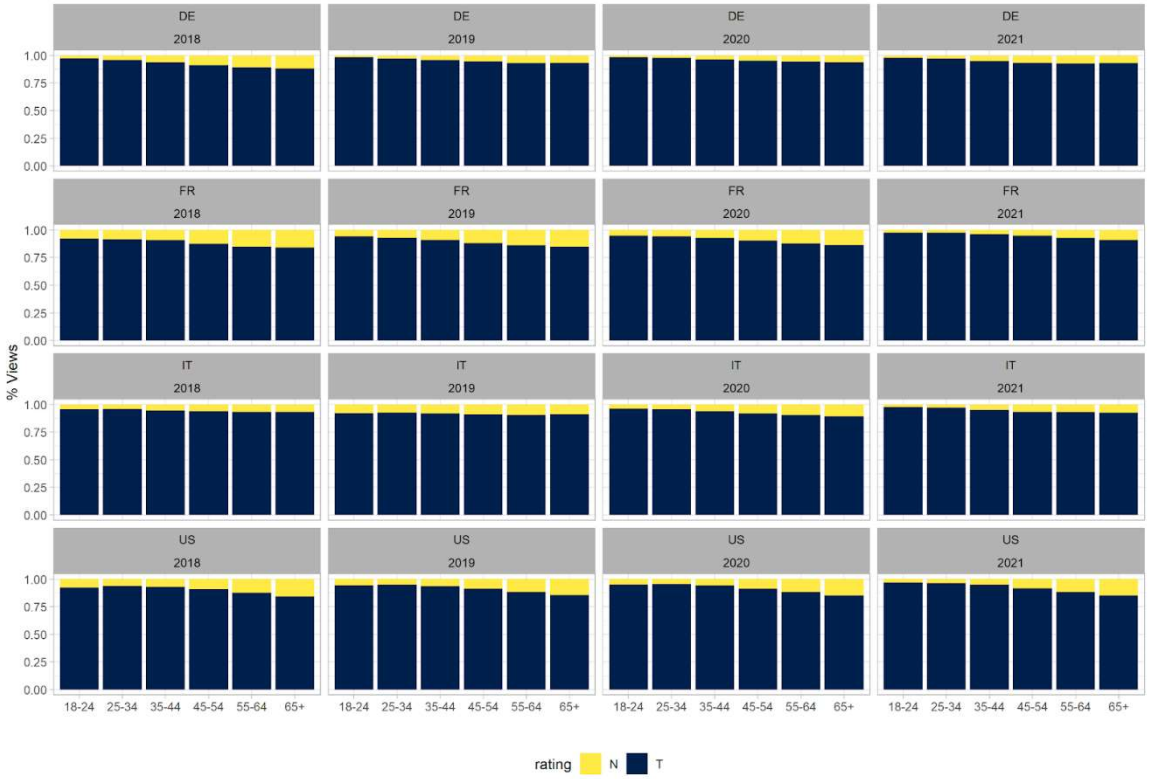

Figure 4: Percentage of Views received by trustworthy and untrustworthy links per age group

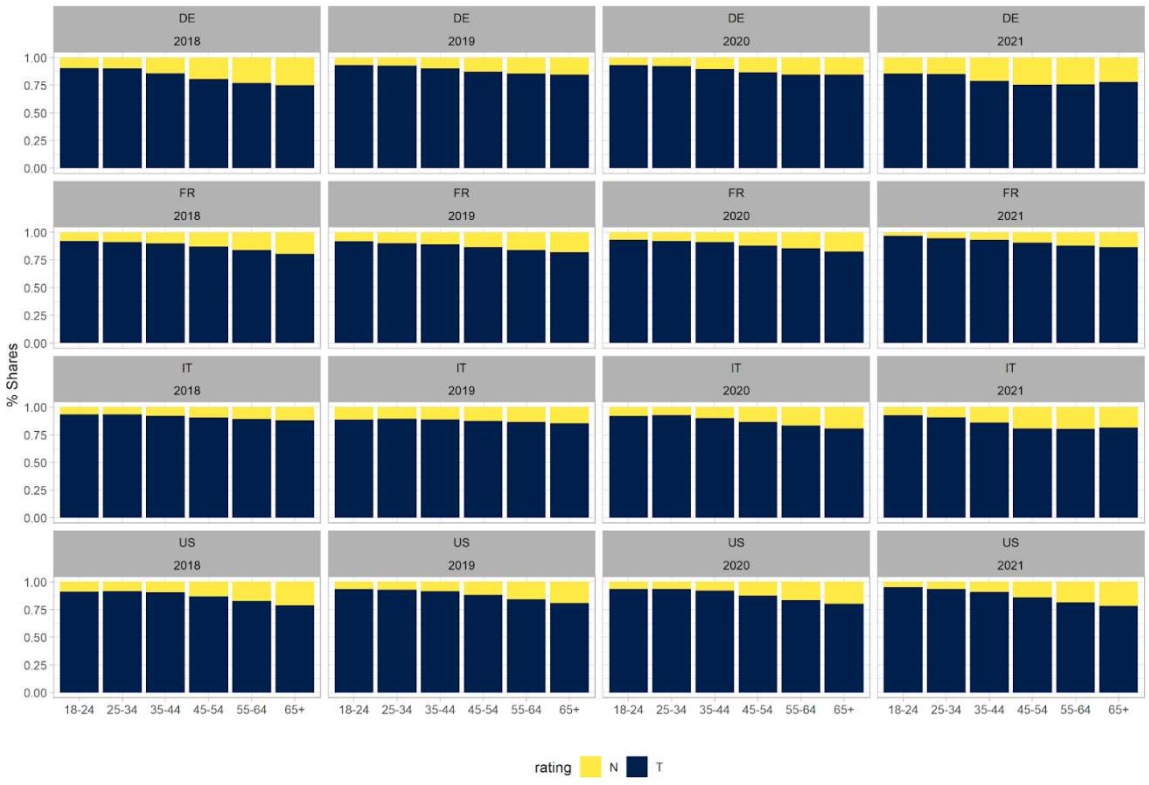

Figure 5: Percentage of Shares received by trustworthy and untrustworthy links per age group