Journal of Quantitative Description: Digital Media 3(2023), 1–45 10.51685/jqd.2023.020

Cracking Open the European Newsfeed

LUCA ROSSI
IT University of Copenhagen, Denmark

FABIO GIGLIETTO
GIADA MARINO
University of Urbino Carlo Bo, Italy

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 the first quantitative assessment of the quality of the information circulating on Facebook among US users. Leveraging an updated version of the same data source — Meta's URL Shares Dataset — and replicating much of the methodology, we quantify the trustworthy and untrustworthy links to external websites that have been shared on Facebook in the period between 2019 and 2022 in three major European countries (Germany, France, and Italy). We observe a clear decline in the number of URLs present in the dataset and an increase in the URLs from untrustworthy domains as a percentage of the total URLs shared in a year. This increase seems to be higher in electoral years (in Germany and in Italy) but it does not translate into an increase of Views received from untrustworthy sources.

Keywords: Facebook, URL Shares Dataset, Trustworthy and untrustworthy news sources.

Author 1: lucr@itu.dk
Author 2: fabio.giglietto@uniurb.it
Author 3: giada.marino@uniurb.it
Date submitted: 2023-08-10

During the last several years, questions about the quality of the information circulating within contemporary social media platforms have been at the center of the research agenda for media scholars. This considerable effort has produced a significant advancement in what we know about problematic online information and the role played by social media platforms in mitigating, supporting, or amplifying the phenomenon (Allcott et al., 2019).

Nevertheless, a significant number of current studies fail to address the critical issue of evaluating the effects of exposure to problematic information (Benkler, 2019). Although assessing the impact of media exposure is already a challenging task, it is particularly complicated when it comes to social media platforms due to the difficulties in measuring exposure. Exposure data (views, reach) are rarely disclosed by digital platforms and are often estimated by scholars using interactions as a proxy. Alternatively, exposure can be assessed using self-reported data or by "passive tracking" the browsing behavior or post-exposure of a sample of the users recruited to participate in the study. The first approach is prone to biases worsened by the high media fragmentation that characterizes the current ecosystem. The second is expensive to set up and the data quality is often affected by the common users' practices of using multiple devices (e.g. mobile phones, tablets, computers) that are rarely all trackable. As a result, the most convincing scientific evidence currently available is specific to certain countries and stems from peculiar collaborations between social media companies and scholars based in the US (González-Bailón et al., 2023; A. M. Guess et al., 2023).

In this paper, we combine Facebook's URL Shares Dataset (Messing et al., 2020) and NewsGuard ratings to estimate and compare the Shares and Views of high-credibility and low-credibility news sources in the US and in three European countries (France, Germany, and Italy). By doing this, we extend both temporarily and geographically the findings, method, and research design developed by Guess et al. (2021).

State of the art & Research questions

Measuring news exposure and sharing behaviors on social media

Over the past seven years, there has been growing concern about news consumption on social media, particularly with regard to the spread of mis/disinformation on these platforms. These concerns were propelled by the unexpected outcomes of the 2016 US election and the Brexit referendum, which raised concerns on the potential consequences of the widespread circulation of problematic content on social media.

In light of the mutated scenario, researchers repurposed older approaches and developed new ones to measure exposure to and peer-to-peer circulation of news. The main aim was to assess the overall health of the online information by measuring the prevalence of mis/disinformation. Survey data is a commonly employed monitoring method to investigate misinformation exposure and sharing practices on social media. The Reuters Digital News Report and similar studies based on surveys (see, for example, (Eady et al., 2023; Martin & Hassan, 2020; Newman et al., 2021) offer valuable perspectives, providing information on users' perceptions of the toxicity of news media environments and their contributions to them.

While this methodological approach yields valuable insights, it comes with several limitations. Given that it relies on self-reported data, the measures obtained are inevitably influenced by users' perception biases. Moreover, the fragmentation of news sources and the practice of encountering content produced by these sources through a social media feed can impede users' ability to identify the source, potentially affecting the quality of the data collected. Lastly, comparative studies across countries are rare due to the costs of time, researchers employed, and expenses that interview representative samples of multiple populations expect.

An exception concerning the comparative perspective is the Reuters annual report that interviews tens of thousands of online news consumers in 46 markets. This comparison opportunity is significant considering that news consumption behaviors may have extreme variability even among countries we would otherwise perceive as culturally similar. For example, while Western countries registered similar percentages of digital media

penetration in 2022, the use of Facebook for news consumption in Europe and the U.S. have significant differences. The 2022 Digital News Report (Newman et al., 2022) reported that Italy and France have a similar percentage of users consuming news on Facebook (respectively, 45 and 39 percent). This percentage is significantly higher than the US (28 percent) and, especially, Germany (17 percent). Similar percentages in Europe and North America concern users' perception of encountering misinformation online about different topics, from COVID-19 to politics and climate change (Newman et al., 2022).

Another way to measure news exposure at individual levels is by monitoring users' digital traces by tracking users' devices. The study by Guess and colleagues (2019), for example, analyzed misinformation-sharing behavior during the 2016 U.S. presidential campaign by tracking Facebook profile data obtained via an authenticated application. The study by Osmundsen and colleagues (2021) asked a YouGov panel of Twitter users the consent to scrape their profile content in a specific timeframe.

This set of methodological techniques allows researchers to only partially circumvent personal perception biases by linking behavioral data of news consumption practices to individual self-reported data. As for survey methodological approaches, sampling outcomes depend on users' willingness to share their private profile data with researchers. Moreover, the sample may be affected by other biases typical of panel data, such as for example misreporting practices or the under-representation of certain social categories due to social desirability biases (Krumpal, 2013).

Possibility of measuring news consumption behaviors at a macro level, i.e., at a societal or significant group level, is, instead, much more challenging. This way of estimating news exposure and sharing may be possible through social media platforms' large datasets made available to researchers by social media companies through specific programs. One of these programs is that of Facebook and Social Science One, which provides vetted researchers access to such datasets (King & Persily, 2020). In its latest iteration (v 10.1), the dataset includes all URLs on Facebook from January 2017 to October 2022 publicly shared by 100 or more users and accompanying interaction metrics like Clicks, Views, and Shares (Messing et al. 2020).

To mathematically safeguard user privacy, noise has been added to action and exposure data using a method called differential privacy. This addition of noise prevents the identification of individual user behaviors, yet aggregated statistical results remain reliable. As argued by Allen and colleagues (2021), this dataset has limitations. The dataset censorship and the differential private noise may affect calculation results. Despite these limits, the URL Shares Dataset is an unprecedented source of observational data on news exposure and sharing behaviors. Based on the URL Shares Dataset, the exploratory analysis on the U.S. of Bailey and colleagues (2021), for example, found that posts containing URLs related to potential mis/disinformation news can attract substantial user engagement. Altay, Nielsen, and Fletcher (2022) analyzed US, UK, France, and German news outlets' digital traces left between 2017 and 2021 to quantify the extent of the "infodemic" related to Covid-19. They observed that online news consumption increased in volume, but trustworthy news outlets benefited the most from this web traffic growth.

Implications of age cohorts on news exposure and sharing

Diverse studies provided evidence that users' age has effects on levels of exposure to misinformation and sharing practices (see, for example, (Allcott et al., 2019; Brashier & Schacter, 2020; Grinberg et al., 2019; Loos & Nijenhuis, 2020). One of the most widespread studies assessing the intersection of belonging to a specific age group and misinformation dissemination is that of Guess and colleagues (2019). Regarding the 2016 U.S. elections, authors found a strong age effect on misinformation-sharing practices. On average, users over 65 shared nearly seven times as much news from problematic domains as the youngest age group. Guess et al. study (2021) that inspired this paper also found an association between exposure and sharing of fake news and age groups. Authors found that political news from low-credibility sources is more prevalent among older Americans than other users. Younger users share the news with clickbait headlines more than other age cohorts.

In line with Guess et al. results (2021) and relying on the same dataset, Bailey, Gregersen, and Roesner (2021) pinpointed that more politically conservative and older

Facebook users are more likely to be exposed to and share those URLs potentially containing mis/disinformation.

Research questions

In light of the considerations mentioned in the previous paragraphs, the present study leverages the analyses conducted by Guess and colleagues (2021) to estimate the prevalence of trustworthy and untrustworthy news source links on Facebook over time and measure exposure to and sharing those links. Moreover, considering that the totality of the study based on the URL Shares Dataset focused on the U.S. context, the present one intends to move its lens from a single country to a comparative perspective and switch the focus to the European context. For this reason, we analyze three European countries, France, Germany, and Italy in the period 2019-2022. Initially, we considered adding the U.S. data to our analysis but methodological considerations (detailed in the *Data and Methods* section) suggested against drawing a comparative perspective between the U.S. data and the European data. Nevertheless, data for the U.S. case has been analyzed and it is presented in Appendix A.

Within this theoretical and empirical framework, we thus formulated the research questions.

Considering opportunities and limitations of URL Shares Dataset for estimating the prevalence of links to trustworthy and untrustworthy news sources, we asked: What is the prevalence of trustworthy and untrustworthy news sources links circulated on Facebook between 2019 and 2022 in France, Germany, and Italy? (RQ1)

Regarding trustworthy and untrustworthy news exposure and sharing activity on Facebook, we asked: Is the total number of *Views* and *Shares* of trustworthy news articles on Facebook greater than the number of *Views* and *Shares* of untrustworthy news? (RQ2) Are there significant differences in *Views* and *Shares* of trustworthy and untrustworthy news source links circulated on Facebook over time in France, Germany, and Italy? (RQ2a)

In light of the considerations about the association of specific age groups and misinformation exposure and sharing, we asked: Are there any significant differences in the way the different age cohorts are exposed to (*Views*) and have shared (*Shares*) low-quality content on Facebook over time in France, Germany, and Italy? (RQ3)

Data and Methods

URL Shares Dataset

To measure the exposure to news stories from trustworthy and untrustworthy news sources on Facebook, we relied on the URL Shares Dataset (Messing et al., 2020) provided by Meta. Unlike other Facebook data sources, this dataset provides a set of user-centered metrics (including exposure) concerning the URLs publicly shared on the platform at least 100 times. The demographics of users who viewed and otherwise interacted with an URL are broken down by gender, age classes, and - only for U.S. users - the user's estimated political leaning. The dataset is usually updated two times a year, and its latest version (v 10.1) covers a period that goes from January 2017 to October 2022. Data is aggregated by month. Repeated exposure or interactions on the same URL by the same user are only counted ones. In other words, the dataset's figures can be read as the number of unique users exposed to or interacted with a certain link.

Given this peculiarity, the metrics presented are secured by differential privacy (D'Orazio et al., 2015; Wood et al., 2018). Differential privacy is a privacy-preserving technique that adds noise to the data to avoid user re-identification. The added noise's impact decreases as the URL aggregations' size increases. A different amount of noise was applied to each metric. 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. At this level of aggregation, the impact of the noise is expected to be negligible (Evans & King, 2023). To answer Research Question 3, we further subdivided our dataset into age categories. To account for this effect, we include an estimate of the noise-induced error in our results. We do so by calculating the confidence

interval around the reported value as defined by Guess at al. (2021) and described in the Equation1. We specifically focus on views (to measure exposure) and shares.

$$\widehat{x_c} \pm 1.96 \times \sqrt{\sum_{r \in c} \sigma^2} \quad (1)$$

Equation 1: Confidence interval for counts in the URL Shares Dataset as defined by Guess et al. 2021. The value of σ = 14 for shares and σ = 2, 228 for views is detailed in Messing et al. 2020.

NewsGuard ratings

To assess the trustworthiness of domains, we utilized the findings of NewsGuard. NewsGuard is a private company that offers various services, including assigning a score ranging from 0 to 100 to each news domain they have evaluated, reflecting their adherence to a set of journalistic standards. A team of specialized journalists investigates each domain, and the scores are periodically reviewed. A total score over 60 indicates a trustworthy news source (T). News sources scoring lower than 60 are labeled untrustworthy (N)[1]. For this study, we used NewsGuard scores dated 31 October 2022 which contain 8,506 unique domains rated.

Given the centrality of NewsGuard data to the current analysis, it is necessary to explore some of its characteristics and assess its overall impact on the results. We will achieve this by addressing three main concerns: 1) the comparability of NewsGuard data with other datasets of low-quality news sources and coverage of European sources; 2) the temporal evolution of NewsGuard scores; 3) the impact of NewsGuard data on the observed results.

Comparing NewsGuard data with alternative datasets of low-quality news sources is not straightforward. NewsGuard adopts a structured approach to news sources

[1] As of this writing, NewsGuard has broadened its categories to five, ranging from High Credibility to Proceed with Maximum Caution. However, the score of 60 remains the benchmark for determining credible sources

monitoring - thus mapping news sources for a specific country first and then evaluating their trustworthiness. This has two major consequences: it produces a clear mapping between the news source and the country and, second, it provides a mapping that does not focus only on spreaders of problematic information but includes a variety of actors (platforms, satire websites, etc.). In order to assess the coverage and possible bias of NewsGuard data we compare it with the collection of domains that spread unreliable information compiled by Jana Lasse and Valenti Rupp [2]available online (LR list from now on). This is a cleaned and consolidated collection of various smaller lists (FactCheck, 2018; Fletcher et al., 2018; LaCapria, 2016; Media Bias / Fact Check, 2015; PolitiFact Staff, 2017; Silverman et al., 2017; Zimdars, 2016). Some of those are based on reporting and assessment from fact-checkers and have some geographical focus (e.g., Bufale.net has a long list of Italian domains) while others have more general approaches that aim at providing maps of national media ecosystems (e.g., Media Bias / Fact Check).

Table 1 presents a comparison between the number of trustworthy and untrustworthy sources in the NewsGuard data, the LR list, the Media Bias / Fact Check (MBF) data, and the data by Grinberg et al. (2019). It is worth noting that the MBF and Grinberg et al. (2019) data are subsets of the LR data.

Table 1. Domains rated as Trustworthy/Untrustworthy (or equivalent label) in various datasets.

Rating	NG	LR	MBF	Grindberg et al.
Trustworthy	5089 (60%)	2597 (45%)	2012 (68%)	-
Untrustworthy	3327 (39%)	2170 (55%)	935 (31%)	372 (100%)
Platform	61 (0.5%)	-	-	-
Satire	54(0.5%)	-	-	-

It can be noted that, on the one hand, NewsGuard offers the largest coverage of news sources while maintaining a balance between trustworthy and untrustworthy sources.

[2] https://github.com/JanaLasser/misinformation_domains

It should also be observed that approaches aimed at mapping the media ecosystem, such as the NewsGuard and the Media Bias/FactCheck datasets, offer comparable results. Additionally, it should be noted that NewsGuard and the LR list share 1,761 domains (38% rated as untrustworthy and 60% as trustworthy).

When it comes to national coverage, assessing the datasets is challenging, as none of them, except for NewsGuard, allow for country-specific analysis. A rough estimate of the top-level domains covered by MBF, the only source attempting to map the news ecosystem with international coverage, indicates a minor presence of Italian, German, or French top-level domains (7 .de; 6 .it; 5 .fr).

To gain further insights, we analyzed NewsGuard historical data. This allowed us to map how different types of trustworthy and untrustworthy sources have been added to the dataset since the second half of 2019 (Figure 1).

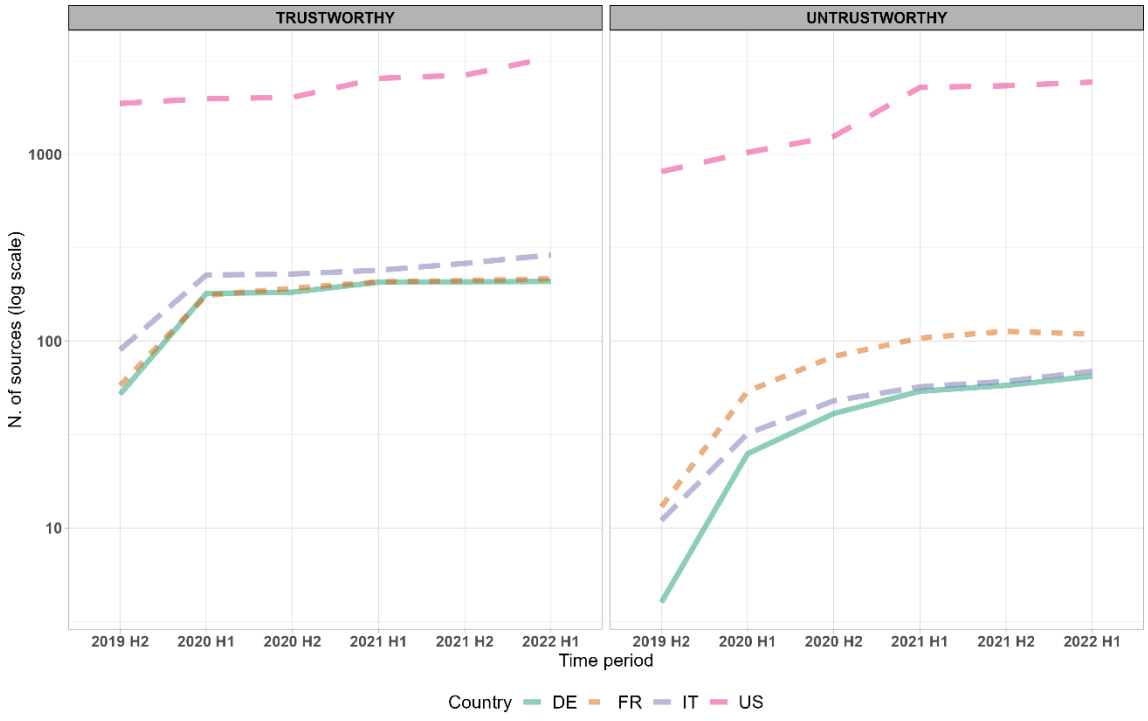

Figure 1. Number of domains in the NewsGuard data.

Figure 1 allows us to observe a steady and comparable rate of adding European sources. However, the number of European sources, as well as the overall number, is considerably lower than what is observed in the U.S. data. Notably, there is a visible difference in the percentage of untrustworthy sources added to the data. In 2022, the data used for this analysis shows that untrustworthy sources represent 42% of the sources in the U.S., while at the EU level, it varies from 19% in Italy to 33% in France. This suggests that a larger presence of domains labeled as untrustworthy in the NewsGuard data could lead to increased detection of URLs shared by untrustworthy sources in the URL Shares Dataset.

To investigate this possibility further, we calculated, for each country, the correlation between the percentage of untrustworthy domains present in the NewsGuard data and the percentage of untrustworthy URLs (the ratio between links belonging to untrustworthy domains and links from all domains) observed in the URL Shares Dataset for each year (Figure 3). The overall correlation is inconclusive and not significant, but a visual examination of Figure 3 reveals a different dynamic in different countries. In the U.S. data, there seems to be a significant and positive correlation, while in the European data, there are cases where an increase in the percentage of untrustworthy domains corresponds to a decrease in the percentage of observed URLs (e.g., DE 2022, FR 2021, FR 2022). Considering these elements, we decided to focus the comparative analysis only on the European data. The U.S. data, which shows a larger percentage of untrustworthy sources and stronger signs of a positive correlation between the percentage of untrustworthy sources in the NewsGuard data and the untrustworthy URLs observed in the URL Shares Dataset, will be reported in Appendix A.

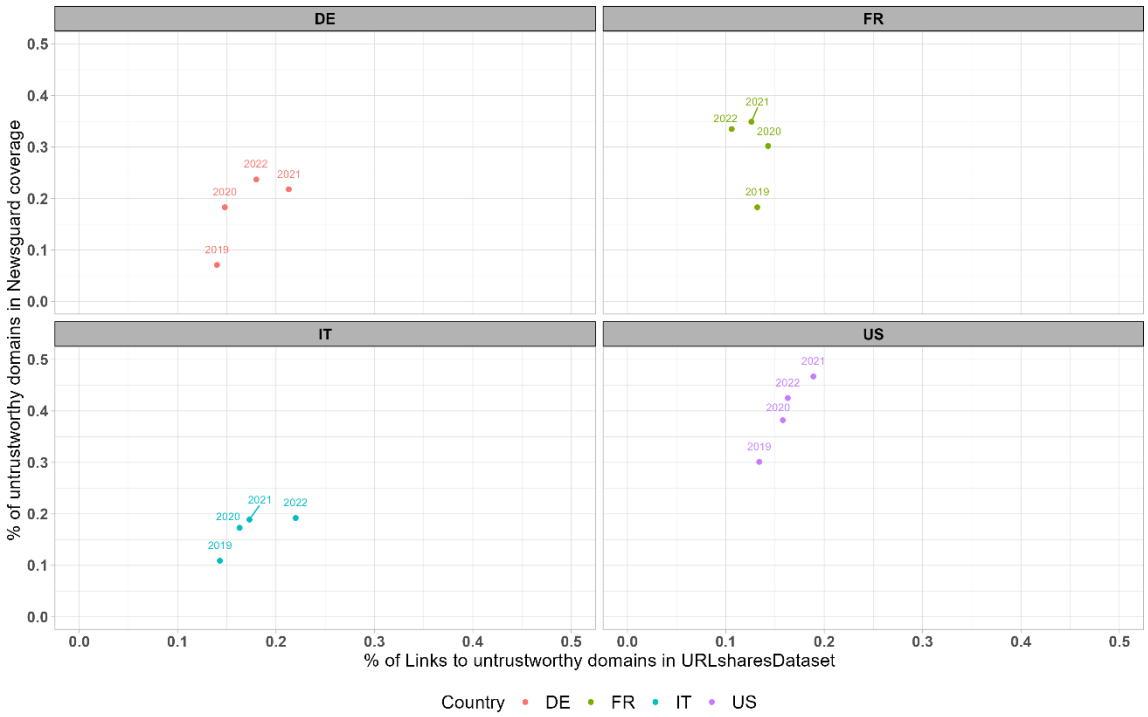

Figure 2. Correlation between the percentage of untrustworthy domains in the NewsGuard data and the percentage of URLs to untrustworthy domains in the URL Shares Dataset.

A final concern regarding the use of longitudinal NewsGuard ratings relates to possible changes in the ratings of sources. As shown in Figures 1 and 2, the NewsGuard data has been accumulating both trustworthy and untrustworthy sources over the years. Figure 2 shows a possible positive correlation between the number of untrustworthy sources in the dataset and the number of links to untrustworthy domains in the URL Shares Dataset in some countries. This suggests that using the most recent NewsGuard data, which includes the largest number of sources per country, might be beneficial. However, it's important to consider that NewsGuard scores are regularly updated, which means that sources could change their scores and ratings over time, moving between being classified as trustworthy and untrustworthy.

To address this concern, we leveraged the historical data provided by NewsGuard. Upon analysis, we found that overall, scores changes large enough to shift the source category are extremely rare events, affecting only 0.01% of the domains. Table 2 presents the percentage of domains that have experienced changes in their ratings over the period from 2019 to 2022 for the three European countries and the U.S. Additionally, Appendix B provides a visualization of all the domains that have seen their rating changed during the analyzed period.

Considering the minimal occurrence of rating changes, this concern is mitigated, and it supports the use of the most recent NewsGuard data in our analysis. The data's consistency and reliability over time make it a valuable resource for understanding and comparing the trustworthiness of sources across countries and regions.

Table 2. Domains with change in "rating"in the NewsGuard dataset.[3]

Country	% of domains with changed rating
DE	1.05
FR	2.67
IT	1.94
US	1.75

Given the overall stability in ratings and the evident impact of the number of domains, we decided to employ the ratings of the domains rated by NewsGuard in October 2022 retrospectively to the entire period.

Additional details

[3] Change in rating: T > N or N > T. Timeframe: 2019 and 2022

To query the URL Shares Dataset for RQ1, we sought the raw monthly count of URLs from each NewsGuard domain that were primarily shared by users from either Italy, Germany, France, or the United States. The country where a URL has been mainly shared in part of the URL Shares Dataset and is represented by the attribute called "top_country". To align with the time period covered by NewsGuard historical data and our ratings, we limited this query from February 1st, 2019, to October 31st, 2022.

For the remaining RQs, as the goal was to replicate the results of a previous US-based study for three European countries, and given that the authors' code is publicly available (Aslett, 2022), we designed our data pull and analysis processes to closely resemble the original design. Consequently, we collected Views and Shares numbers (or the number of unique users who viewed and shared) for all the URLs published between 2018 and 2020 belonging to the domains rated by NewsGuard and predominantly shared by users from the countries under investigation. Additionally, for each URL, we requested data broken down by month and user age class (18-24, 25-34, 35-44, 45-54, 55-64, 65+).

Unlike the original studies, we decided to include all the views and shares obtained by URLs published between 2019 and 2022, regardless of the year they were initially published. The rationale for the original studies was to measure the prevalence of URLs published in a specific year. However, this design did not account for views and shares recorded by a URL in subsequent years and, consequently, penalizes URLs published during the last months of the year. To overcome this limit and considering the frequent resurgence of old content, we opted to take this possibility into account. This approach applies to both trustworthy and untrustworthy news sources.

Results

Prior to addressing the distinct research questions, it is pertinent to scrutinize the aggregate count of URLs originating from both Trustworthy and Untrustworthy domains, as delineated by NewsGuard, disseminated across the four designated nations during the specified time interval. From 2019 to 2022, a cumulative total of 1,981,469 URLs were prevalently seen by Facebook's users across Germany, France, Italy, and the United States.

Figure 3 provides a visual representation of the temporal distribution of URLs within the dataset.

The discrepancies among Germany (population approx. 83 million), France (approx. 65 million), and Italy (approx. 59 million) are noteworthy. Despite the overall population in each country and recent Eurostat data reporting a comparable level of adoption of digital technologies for social networking (van Kessel et al., 2022), the three countries exhibit a significantly different volume of links circulating on Facebook, which remains consistent throughout the analysis period. France consistently has the highest number of URLs being shared on Facebook, with Italy following as a close second. On the other hand, Germany has the lowest number of URLs, often comprising half or even less than what is shared in France. This initial observation aligns with findings from the 2022 Digital News Report (Newman et al., 2022) concerning Facebook adoption for news consumption. Furthermore, it is possible to observe a similar declining trend for the three countries, where URLs in the data constantly decreased over time.

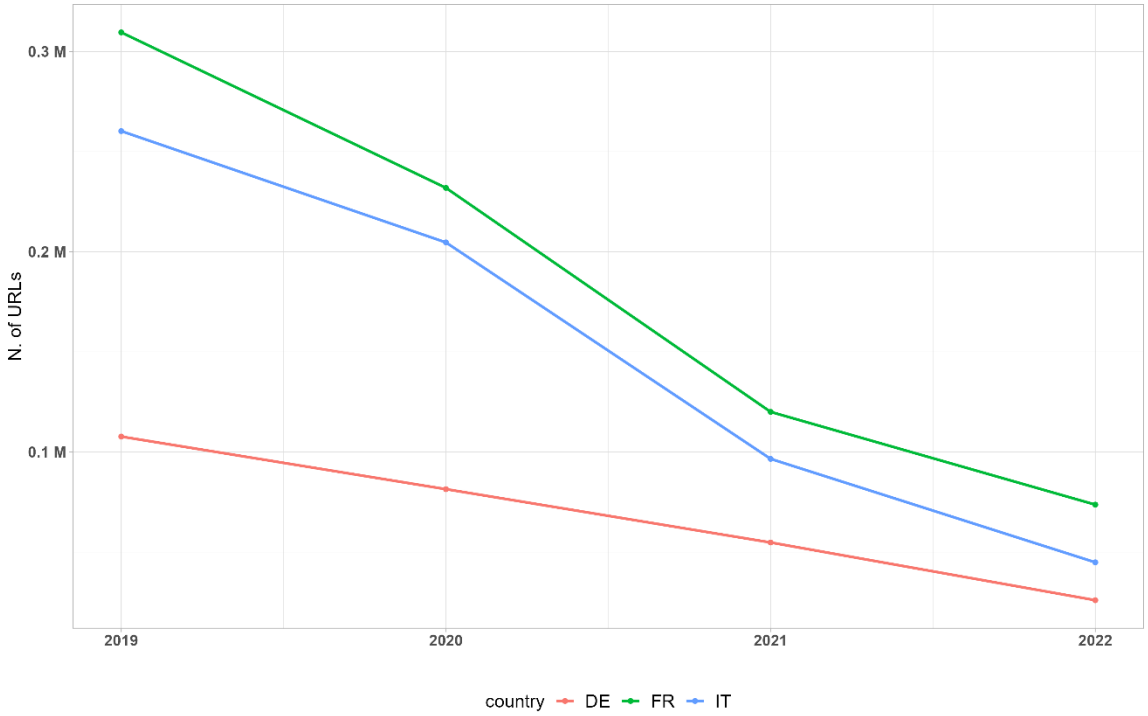

Figure 3. Number of URLs in the URL Shares Dataset over time in France, Germany, and Italy.

Upon analyzing the percentage of URLs from untrustworthy sources, a distinctive pattern becomes apparent (refer to Figure 4). In 2019, all three countries displayed a comparable proportion of URLs from dubious sources: France at 13%, Italy at 14%, and Germany also at 14%. However, their trends diverged in the ensuing years. France's percentage exhibited relative stability, experiencing a modest decline to 11% by 2022. In contrast, Germany witnessed a notable surge in URLs from untrustworthy domains, culminating in 21% by 2021. Similarly, Italy's percentage also escalated, climaxing at 22% in 2022.

It is noteworthy that both Germany and Italy held their general elections in their corresponding peak years. The correspondence between the election period and the peak of URLs from untrustworthy sources could suggest a connection between these two events.

However, in France, we observe the exact opposite: the 2022 elections coincide with the lowest level of URLs from untrustworthy sources. This finding adds complexity to the picture indicating that different factors might be influencing the prevalence of URLs from untrustworthy sources in each country.

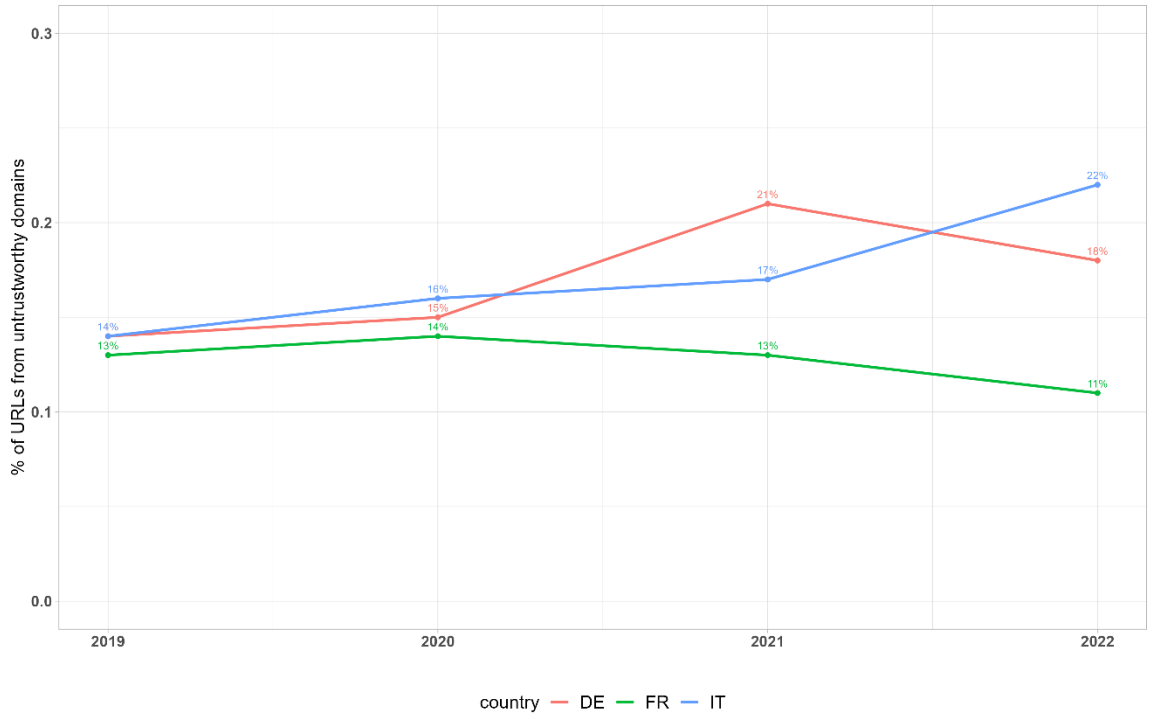

Figure 4. Percentage of URLs from untrustworthy sources over time in France, Germany, and Italy.

Exposure and Sharing of untrustworthy sources

Exposure to untrustworthy sources

To address Research Questions 2 and 2a, we scrutinize both the exposure to and dissemination of links deemed untrustworthy over the span of four years. The analytical outcomes underscore temporal disparities among the nations. Figure 5a delineates the absolute count of views garnered by untrustworthy sources from 2019 to 2022. It's imperative to acknowledge that the lines on the graph signify the margin of error attributable to the implementation of differential privacy measures. However, at this granularity, such error remains inconsequential. Conversely, Figure 5b vividly depicts the proportional views amassed by links emanating from these untrustworthy domains.

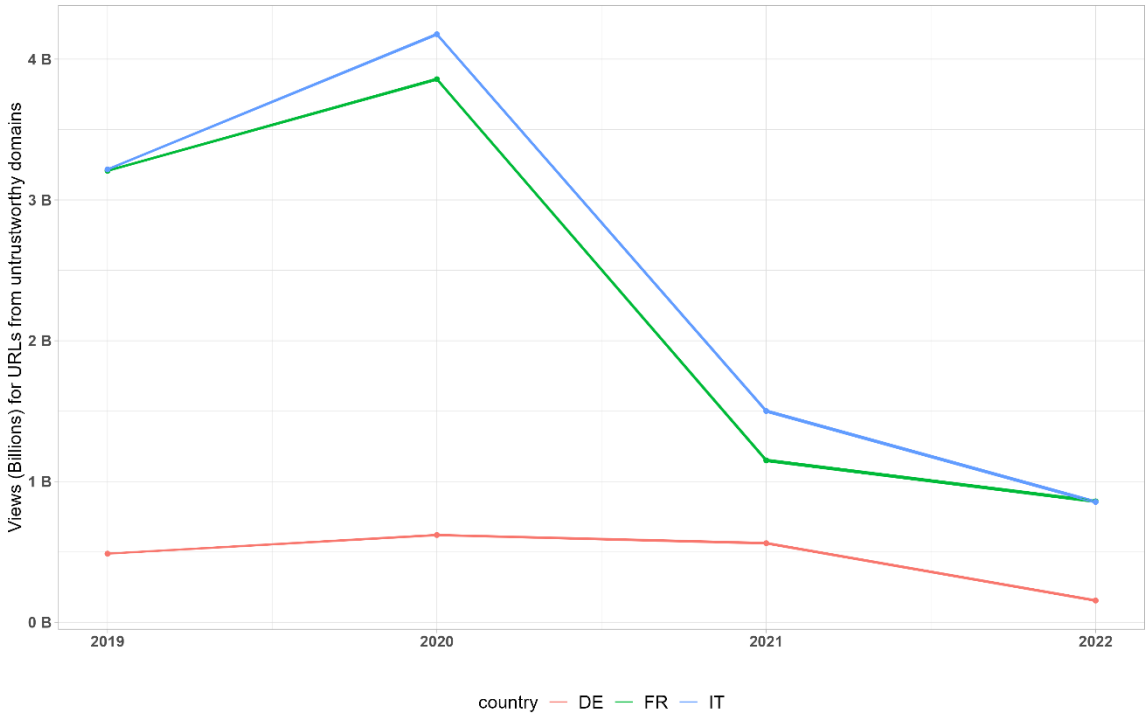

Figure 5a. Number of Views for URLs from untrustworthy sources.

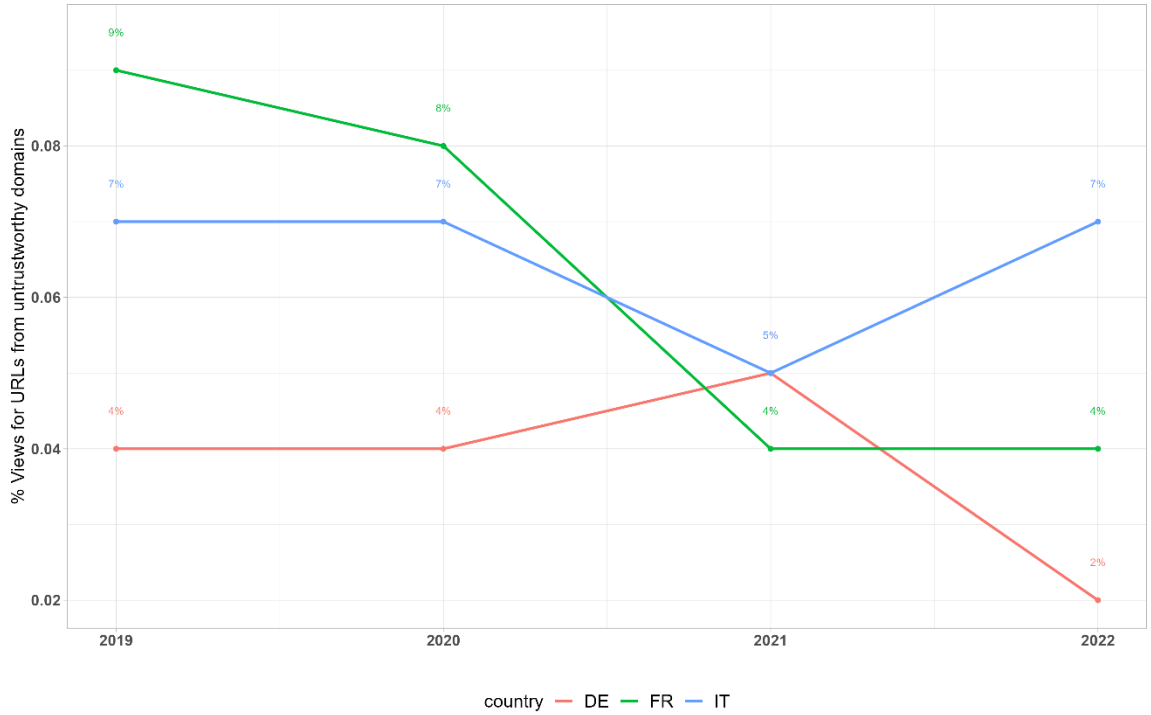

Figure 5b. Percentage of Views for URLs from untrustworthy sources.

The absolute number of *Views* for URLs from untrustworthy sources exhibits a significant decline after it peaked in 2020 for France and Italy, with the exception of Germany, where the number of views remains substantially stable. This trend is remarkably similar to the trend observed for the absolute number of URLs in the dataset.

However, looking at the percentage of *Views* for URLs from untrustworthy domains in Figure 5b, this observation is not confirmed. As depicted in Figure 5b, the percentage of *Views* from untrustworthy sources shows either a small decline for France and Germany or overall stability for Italy. This contrasts with the percentage of untrustworthy sources' URLs in Figure 4 which shows a clear growth for Italy and Germany and a small decline for France. While in the case of Germany and Italy, one could see some effect of the electoral cycle (Germany reaching its highest percentage of *Views* in 2021 and Italy

reversing a downward trend in 2022), these effects appear to be relatively small when compared with the increase in the percentage of URLs from untrustworthy sources observed in the corresponding years.

Sharing of untrustworthy sources

Shares for URLs from untrustworthy sources (Figures 6a and 6b) provide the other half of the story, complementing the information conveyed by the number of Views. They serve as metrics of both circulation and engagement for specific content.

The trends observed for *Shares* are remarkably similar to what we have observed for the number of Views. Overall, the absolute number of *Shares* received from URLs from untrustworthy sources (Figure 6a) has declined, even though Italy and France show a remarkable peak in 2020, and their level in 2021 and 2022 is still noticeably higher than Germany's.

When looking at the percentage of *Shares* obtained by URLs from untrustworthy domains, the data look remarkably similar to the overall percentage of URLs from untrustworthy domains in the dataset, with Germany and Italy reaching 22% and 26% of Shares for untrustworthy domains, respectively, in 2021 and 2022. When we pair this data with the percentage of *Views* received by the URLs from untrustworthy sources - which remains overall stable, if not showing a small decline - we see that the increase in engagement and sharing activity for untrustworthy content does not seem to generate a considerable number of new *Views*.

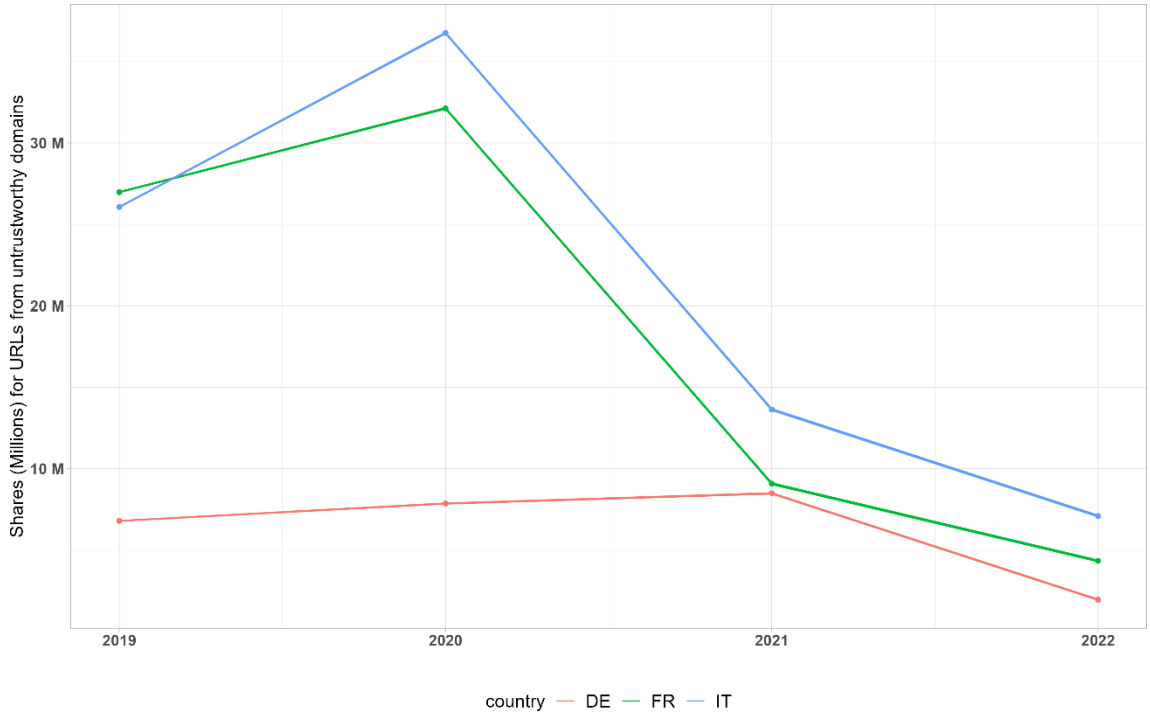

Figure 6a. Number of Shares for URLs from untrustworthy sources.

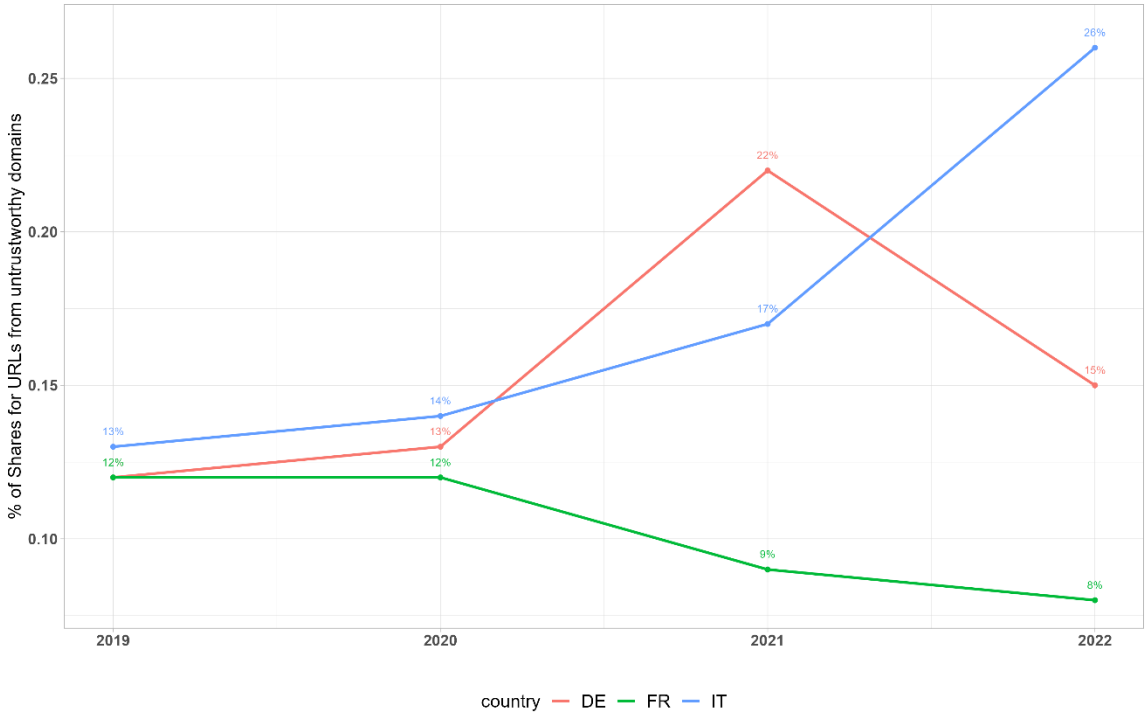

Figure 6b. Percentage of Shares for URLs from untrustworthy sources.

Exposure and Sharing per age group

In this section, we present the absolute numbers (Figure 7a) as well as the percentage (Figure 7b) of *Views* received by URLs from untrustworthy sources divided per age group. The data reveal that exposure to content from untrustworthy sources is more significant among individuals aged 35 to 64. While there are some similarities between the countries - with the age group between 18 to 24 consistently being the least exposed to URLs from untrustworthy sources - some differences are observable.

The age of Italian Facebook users exposed to URLs from untrustworthy sources appears to be higher than that of German or French users. In the Italian case, the three most exposed age groups are those between 35 to 64, whereas, in Germany and France, this

includes the 25 to 34 age group as well. The Italian exception could partially be explained by the country's demographic structure, which shows an older median age (Eurostat, 2023).

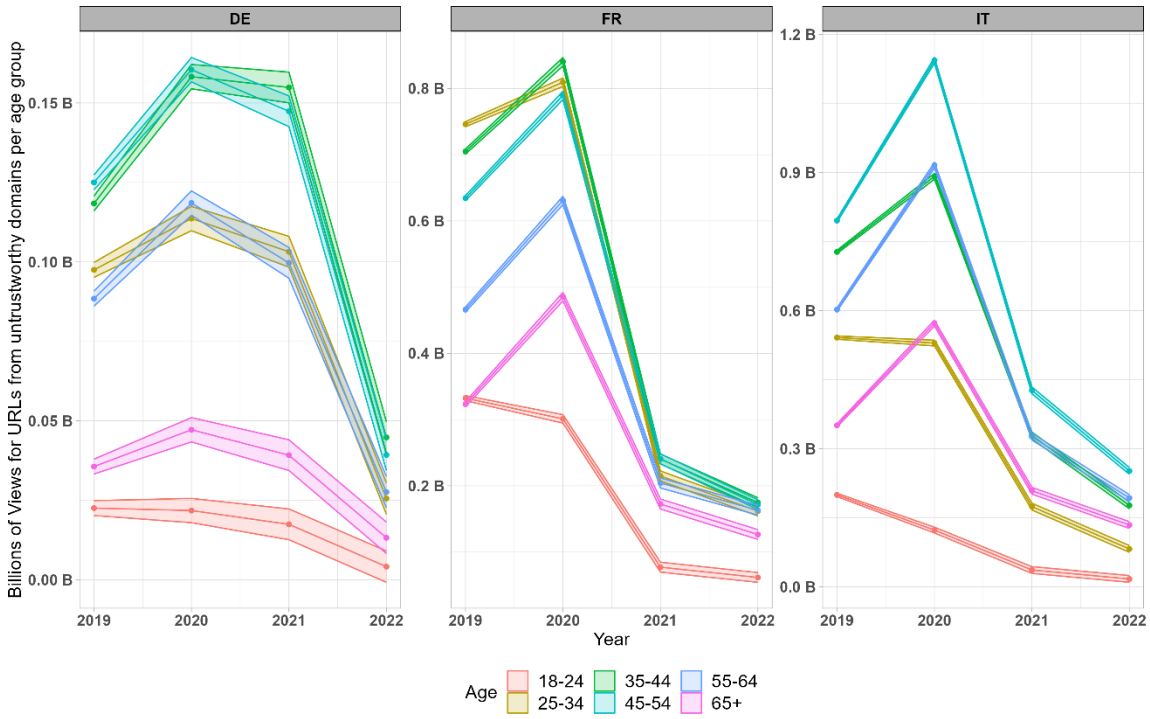

Figure 7a. Number of Views for URLs from untrustworthy sources per age group.[4]

[4] The shade around the line represents the 95% CI calculated as detailed in the Data and Methods section. Please notice that the scale of the Y-axis is different between the three countries.

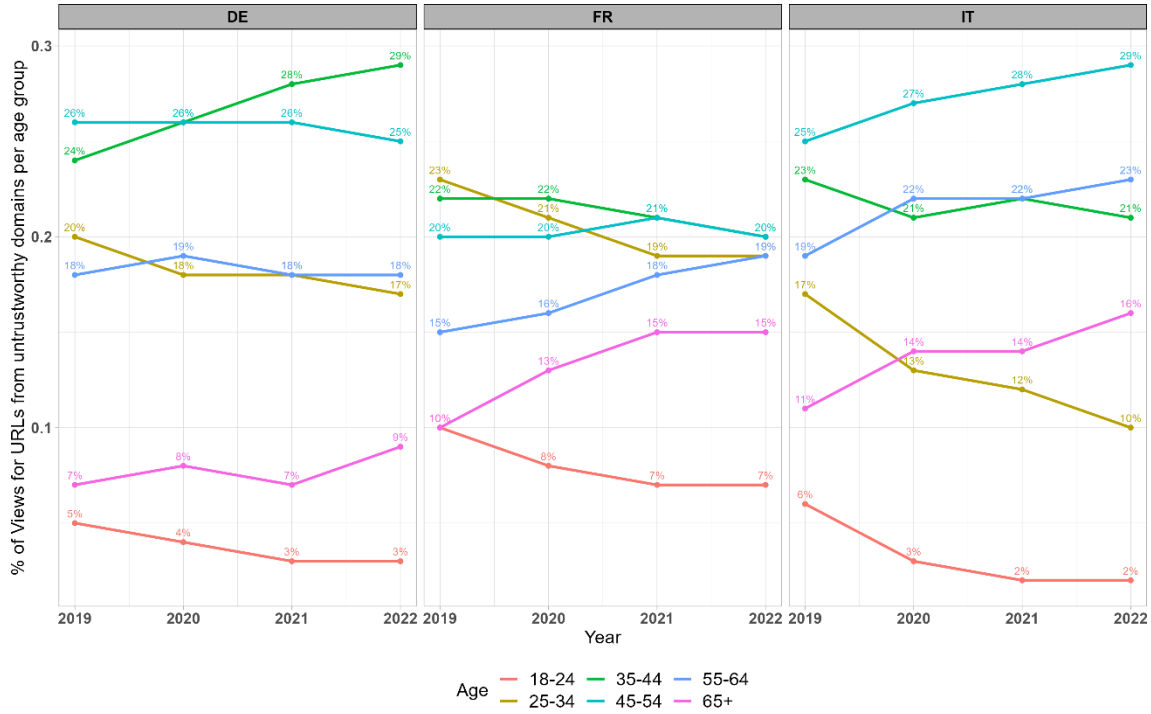

Figure7b. Percentage of Views for URLs from untrustworthy sources per age group[5].

Sharing per age group

The last data we analyze are the *Shares* received from URLs from untrustworthy sources for different age groups. The *Sharing* data reveals a slightly different picture than exposure. *Sharing* of URLs from untrustworthy sources seems to be performed by a population older than the one exposed to it, but there is no clear unique pattern across the three countries.

In Germany, the age group 55 to 64 is responsible for 36% of the *Shares*, in France, the age group older than 65 is the most active with 28% of the *Shares*, while in Italy, it is

[5] The shade around the line represents the 95% CI calculated as detailed in the Data and Methods section. Please notice that the scale of the Y-axis is different between the three countries.

the age group 45-54 with 33% of the *Shares*. With the exception of the Italian case, the age group performing the most *Shares* is older than the age group that is mostly exposed to content from untrustworthy sources.

Nevertheless, both in the case of engagement (*Shares*) and - to a minor extent - in the case of exposure (*Views*), the age groups younger than 34 play a very modest role. Overall, they seem to be less involved in sharing and receiving content from untrustworthy sources compared to the older age groups.

Figure 8a. Number of Shares for URLs from untrustworthy sources per age group.

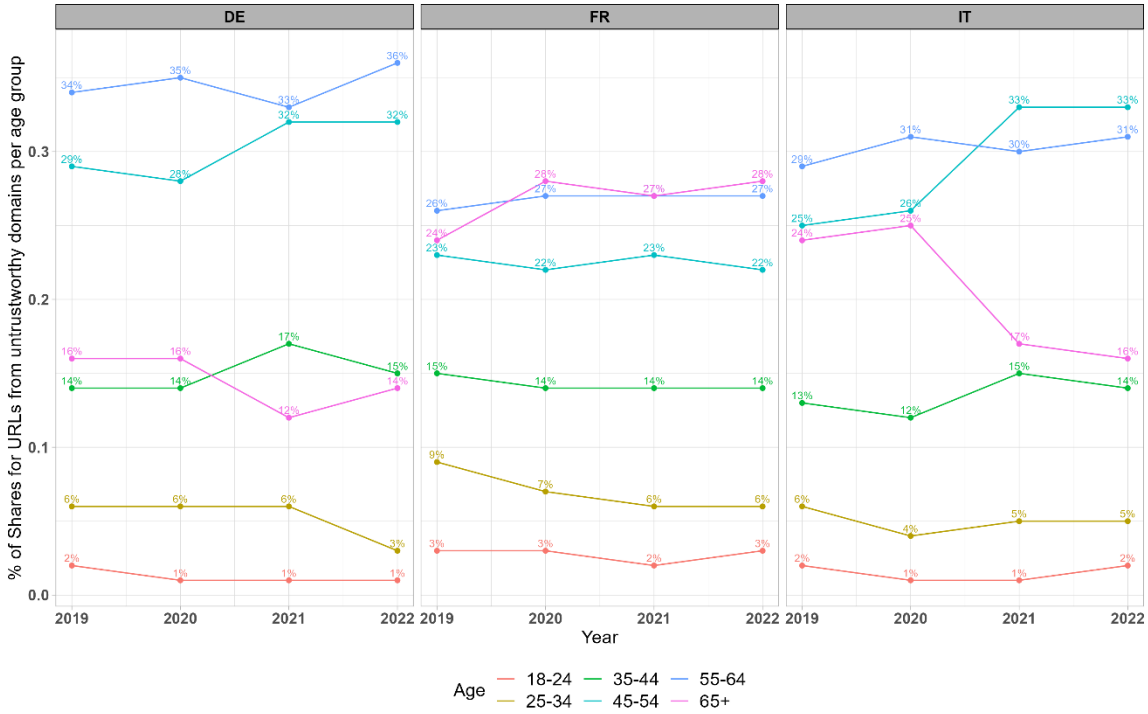

Figure 8b. Percentage of Shares for URLs from untrustworthy sources per age group.

Conclusions

In this study, we have analyzed the presence of URLs by trustworthy and untrustworthy news sources, as well as their *Views* and *Shares* received on Facebook for Germany, France, and Italy in the period from 2019 to 2022. While the paper's aims are mainly descriptive, interesting insights emerge from the data analysis, in particular raising some reflections on the URL Shares Dataset itself, and its opportunities for studying individuals' behaviors on this bulk of observational data as well as its limitations.

A remarkable decline in the number of URLs in the URL Shares Dataset has been observed (Figure 3), starting in 2019 and becoming steeper after 2020, particularly for France and Italy, which consistently have a larger number of URLs and activity on Facebook. The reason for this decline is not clearly identified, and it could be attributed to

distinct causes or a combination of those. This does not seem to be a construct of using the NewsGuard data to filter the news in the URL Shares dataset since NewsGuard has a growing and comparable number of domains analyzed in the same period of time. It may be the result of an organic decline in Facebook's engagement with fewer URLs being posted on the platform (Bell, 2019). Another source of influence could be the specific nature of the URL Shares Dataset, particularly the threshold of 100 public shares that defines the URLs included. This threshold might have affected smaller non-original sources' inclusion in the dataset after Facebook prioritized original news reporting in 2020 (Brown & Levin, 2020). Furthermore, is it possible that the effect of the threshold is magnified by the organic decline in Facebook's overall use and engagement (the overall decrease of active users can hinder the possibility of a URL to pass the threshold).

Although the absolute numbers of *Views* and *Shares* for untrustworthy sources increase from 2019 to 2020, they drop significantly after 2020, especially for France and Italy. Interestingly, these changes in absolute numbers are not fully reflected in the percentages of URLs (Figure 4), *Views* (Figure 5b), and *Shares* (Figure 6b) from untrustworthy sources. The percentage of such URLs declines for France, increases substantially for Germany between 2020 and 2021, and for Italy between 2021 and 2022. A similar trend is visible in the percentage of *Shares* for untrustworthy sources, which reached 22% of the total in Germany in 2021 and 26% in Italy in 2022. The *Views* data, however, show a stable slow decline with small fluctuations for Germany and Italy during the election years. Overall, these data suggest interesting directions for future research both about the observed differences between the actual circulation of URLs from untrustworthy sources (how can we reconcile the high percentage of URLs and *Shares* with a low and stable percentage of *Views*) as well as about the possible French exceptionalism.

Lastly, the data divided per age group show interesting peculiar dynamics. The age structure of Facebook users remains relatively stable between 2019 and 2022, with younger users representing a smaller group in terms of *Views* and *Shares*, while the most active groups are the 35-54 age group for France and Germany, and the 45-64 age group for Italy. Notably, France sees the oldest group in the data (65+) dominating the *Sharing* of URLs

from untrustworthy sources, while in Germany, it is the 55-64 age group, and in Italy, it is the 45-54 age group.

These findings highlight two well-known trends in news consumption and engagement on Facebook. Firstly, the disaffection of younger users for the platform (Auxier & Anderson, 2021). Secondly and in line with what was observed by Guess and colleagues in the United States (2019, 2021), the central role that older generations seem to play in the active dissemination of content from untrustworthy sources. This last aspect is often overlooked in contemporary research and intervention practices aimed at mitigating the effects of exposure to untrustworthy information (Chiluwa & Samoilenko, 2019; Nielsen & Graves, 2017; Shu et al., 2017). The underestimation of the phenomenon is further compounded by fragmentation in the analyses and results obtained when investigating it outside the United States (Chadwick & Vaccari, 2019; Rossini et al., 2021; Valenzuela et al., 2019; Veeriah, 2021), making it even more challenging to hypothesize about the differences observed among the various European countries under analysis.

On a final note, we want to remark on the nature and peculiarities of the URL Shares Dataset. Its specific characteristics have several impactful consequences on researchers' ability to interpret the data fully. The major limitation is the arbitrary threshold of 100 Shares which has proven to be challenging in interpreting trends observable from the data in light of recent phenomena and existing theories. This, coupled with the lack of transparency regarding changes to algorithms supporting content propagation, poses further limitations to researchers' ability to scrutinize the data effectively. While the URLs Shares Dataset represents a unique opportunity to study observational data on user behaviors on Facebook, both from private and public accounts, while also preserving their privacy, such technical decisions should be reconsidered in light of upcoming obligations connected with future Digital Services Act (Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022).

Funding

This work was partially supported by the vera.ai project, which is funded by the European Union through the Horizon Europe program (Grant Agreement ID 101070093).

References

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. Research & Politics, 6(2), 2053168019848554. https://doi.org/10.1177/2053168019848554

Allen, J., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Research note: Examining potential bias in large-scale censored data. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-74

Altay, S., Kleis Nielsen, R., & Fletcher, R. (2022). Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. Journal of Quantitative Description: Digital Media, 2. https://doi.org/10.51685/jqd.2022.020

Aslett, K. (2022). Cracking Open the News Feed: Exploring What U.S. Facebook Users See and Share with Large-Scale Platform Data. https://github.com/kaslett/Cracking_Open_FB

Auxier, B., & Anderson, M. (2021, April 7). Social Media Use in 2021. Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf

Bailey, A., Gregersen, T., & Roesner, F. (2021). Interactions with Potential Mis/Disinformation URLs Among U.S. Users on Facebook, 2017-2019. Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet, 16–26. https://doi.org/10.1145/3473604.3474561

Bell, K. (2019, May 28). Facebook's engagement is sinking with no end in sight. Mashable. https://mashable.com/article/facebook-use-declining-emarketer#:~:text=Engagement%20with%20Facebook%20is%20set%20to%20decline%20or,U.S.%20users%20in%202018%2C%20according%20to%20the%20firm

Benkler, Y. (2019). Cautionary notes on disinformation and the origins of distrust. Social Science Research Counci. https://doi.org/10.35650/md.2004.d.2019

Brashier, N. M., & Schacter, D. L. (2020). Aging in an Era of Fake News. Current Directions in Psychological Science, 29(3), 316–323. https://doi.org/10.1177/0963721420915872

Brown, C., & Levin, J. (2020, June 30). Prioritizing Original News Reporting on Facebook. Meta. https://about.fb.com/news/2020/06/prioritizing-original-news-reporting-on-facebook/

Chadwick, A., & Vaccari, C. (2019). News sharing on UK social media: misinformation, disinformation, and correction. https://repository.lboro.ac.uk/articles/report/News_sharing_on_UK_social_media_misinformation_disinformation_and_correction/9471269

Chiluwa, I. E., & Samoilenko, S. A. (2019). Handbook of Research on Deception, Fake News, and Misinformation Online. IGI Global.

D'Orazio, V., Honaker, J., & King, G. (2015). Differential Privacy for Social Science Inference. In Sloan Foundation Economics. https://doi.org/10.2139/ssrn.2676160

Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. Nature Communications, 14(62), 1–11. https://doi.org/10.1038/s41467-022-35576-9

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October

2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act), European Parliament, October 19, 2022, Official Journal of the European Union, https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Eurostat. (2023, February). Population structure and ageing [2022]. Ec.europa.eu. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing#Median_age_is_highest_in_Italy_and_lowest_in_Cyprus

Evans, G., & King, G. (2023). Statistically Valid Inferences from Differentially Private Data Releases, with Application to the Facebook URLs Dataset. Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association, 31(1), 1–21. https://doi.org/10.1017/pan.2022.1

FactCheck. (2018, November 12). Misinformation Directory. FactCheck.org. https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of "fake news" and online disinformation in Europe. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-reach-fake-news-and-online-disinformation-europe

González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., … Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. Science, 381(6656), 392–398. https://doi.org/10.1126/science.ade7138

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. Science, 363(6425), 374–378. https://doi.org/10.1126/science.aau2706

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. https://journalqd.org/article/view/2586

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. Science, 381(6656), 404–408. https://doi.org/10.1126/science.add8424

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

King, G., & Persily, N. (2020). A New Model for Industry–Academic Partnerships. PS, Political Science & Politics, 53(4), 703–709. https://doi.org/10.1017/S1049096519001021

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. Quality & Quantity, 47, 2025–2047. https://doi.org/DOI 10.1007/s11135-011-9640-9

LaCapria, K. (2016, January 14). Snopes' Field Guide to Fake News Sites and Hoax Purveyors. Snopes. https://www.snopes.com/news/2016/01/14/fake-news-sites/

Loos, E., & Nijenhuis, J. (2020). Consuming Fake News: A Matter of Age? The Perception of Political Fake News Stories in Facebook Ads. Human Aspects of IT for the Aged Population. Technology and Society, 69–88. https://doi.org/10.1007/978-3-030-50232-4_6

Martin, J. D., & Hassan, F. (2020). News Media Credibility Ratings and Perceptions of Online Fake News Exposure in Five Countries. Journalism Studies, 21(16), 2215–2233. https://doi.org/10.1080/1461670X.2020.1827970

Media Bias / Fact Check. (2015). About. Media Bias / Fact Check. https://mediabiasfactcheck.com/about/

Messing, S., DeGregorio, C., Hillenbrand, B., King, G., Mahanti, S., Mukerjee, Z., Nayak, C., Persily, N., State, Bogdan, & Wilkins, A. (2020). Facebook privacy-protected full URLs data set [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/TDOAPG

Newman, N., Fletcher, R., Kirsten, R. C. T., & Kleis Nielsen, R. (2022). Reuters Institute Digital News Report 2022. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2021). Reuters Institute Digital News Report 2021. https://papers.ssrn.com/abstract=3873260

Nielsen, R., & Graves, L. (2017). "News you don't believe": Audience perspectives on fake news. Reuters Institute for the Study of Journalism. https://ora.ox.ac.uk/objects/uuid:6eff4d14-bc72-404d-b78a-4c2573459ab8

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. The American Political Science Review, 115(3), 999–1015. https://doi.org/10.1017/S0003055421000290

PolitiFact Staff. (2017, April 20). PolitiFact's guide to fake news websites and what they peddle. Politifact. https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/

Rossini, P., Baptista, É. A., Veiga de Oliveira, V., & Stromer-Galley, J. (2021). Digital media landscape in Brazil: Political (mis)information and participation on Facebook and WhatsApp. Journal of Quantitative Description: Digital Media , 1, 1–27. https://doi.org/10.51685/jqd.2021.015

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social

Media: A Data Mining Perspective. SIGKDD Explor. Newsl., 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

Silverman, C., Lytvynenko, J., Vo, L. T., & Singer-Vine, J. (2017, August 8). Inside the partisan fight for your news feed. BuzzFeed.News. https://www.buzzfeednews.com/article/craigsilverman/inside-the-partisan-fight-for-your-news-feed

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation. Digital Journalism, 7(6), 802–823. https://doi.org/10.1080/21670811.2019.1623701

van Kessel, R., Wong, B. L. H., Rubinić, I., O'Nuallain, E., & Czabanowska, K. (2022). Is Europe prepared to go digital? Making the case for developing digital capacity: An exploratory analysis of Eurostat survey data. PLOS Digital Health, 1(2), 1–15. https://doi.org/10.1371/journal.pdig.0000013

Veeriah, J. (2021). Young adults' ability to detect fake news and their new media literacy level in the wake of the covid-19 pandemic. Journal of Content, Community and Communication, 13(7), 372–383. https://doi.org/10.31620/JCCC.06.21/31

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D. R., Steinke, T., & Vadhan, S. (2018). Differential privacy: A primer for a non-technical audience. Vand. J. Ent. & Tech. L., 21, 209. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/vanep21&section=8&casa_token=xfB19otGGtMAAAAA:135Hn_NloAQeWaRAHQM7eHXKQPscGiE9LV9o0RfawcFIRpZc7z77sY05jghgxfY2nZkWQNAKHg

Zimdars, M. (2016, November 18). My "fake news list" went viral. But made-up stories are only part of the problem. The Washington Post. https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/

Appendix A

Data for URLs with top_country US.

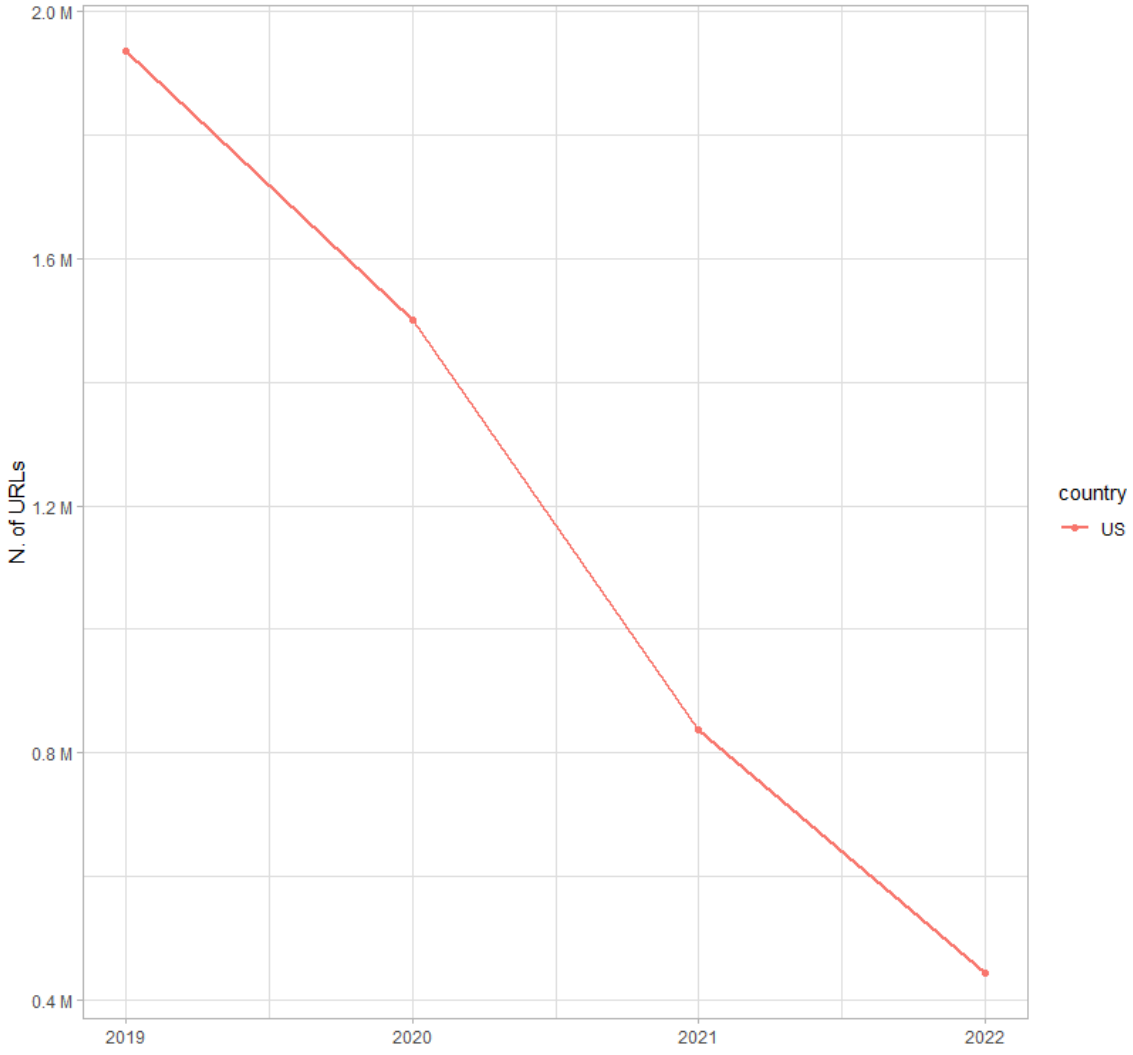

Figure 1. Number of URLs in the dataset

Figure 2. Percentage of URLs from untrustworthy sources

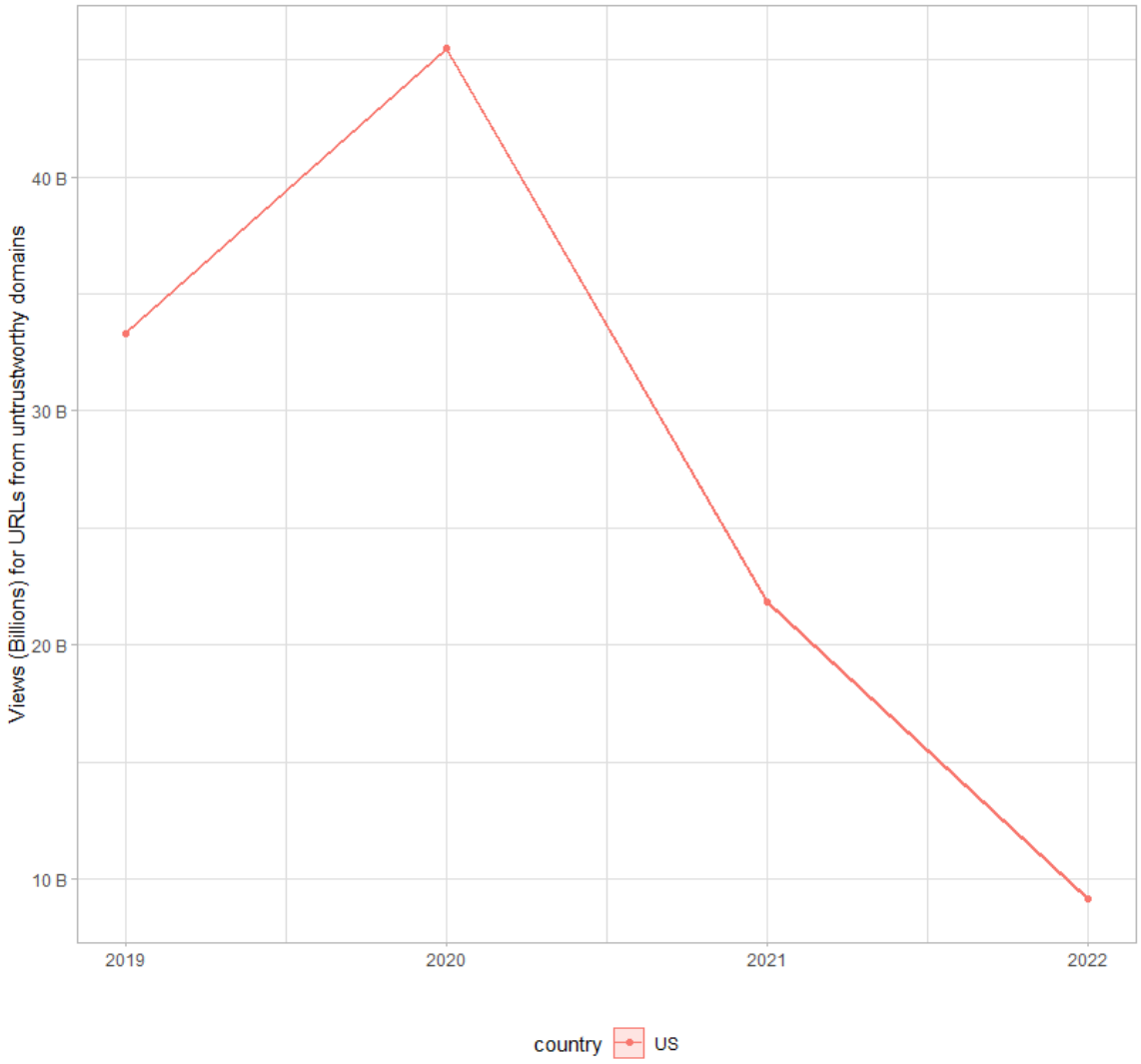

Figure 3. Views for URLs from untrustworthy sources

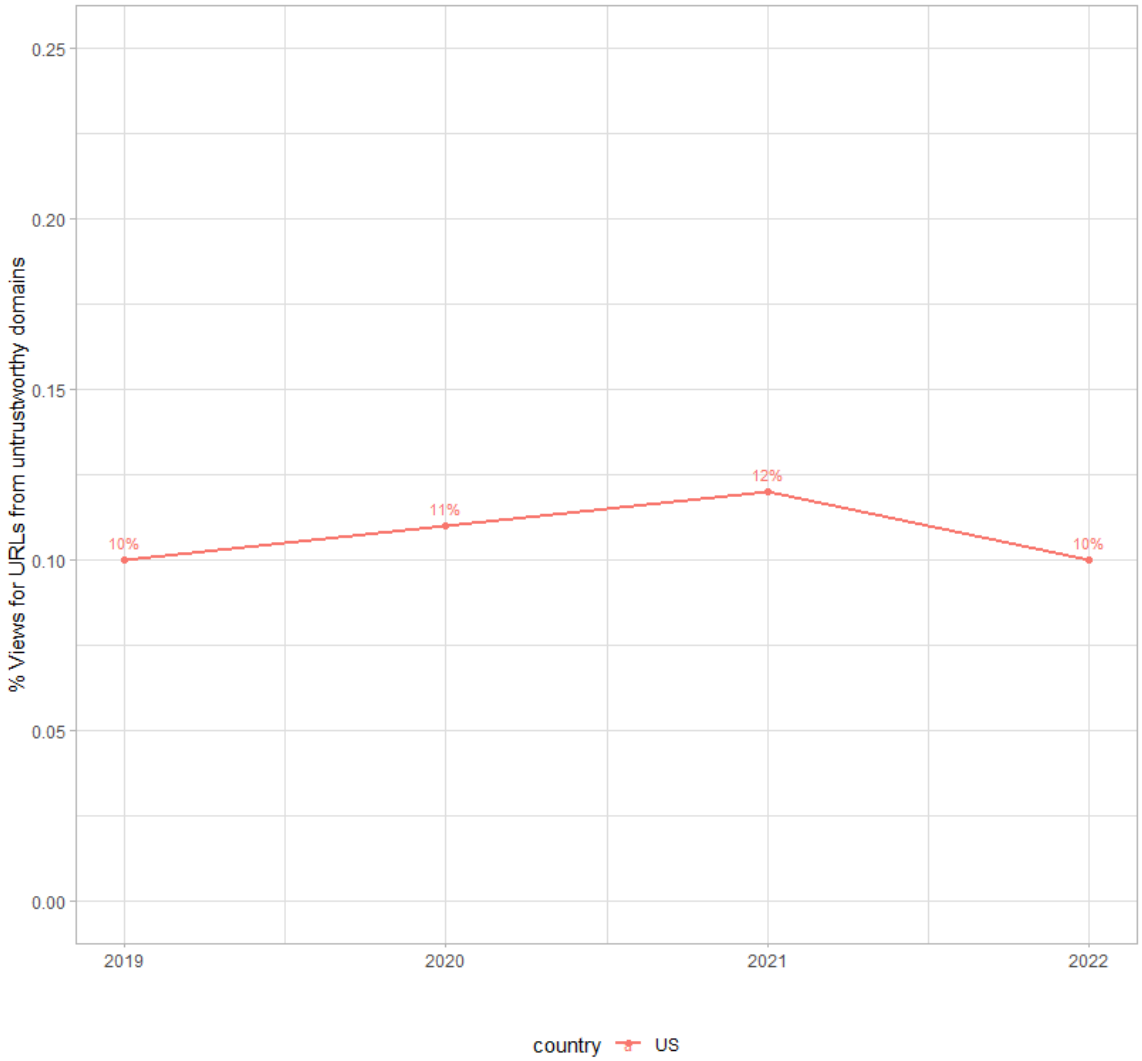

Figure 4. Percentage of *Views* for URLs from untrustworthy sources

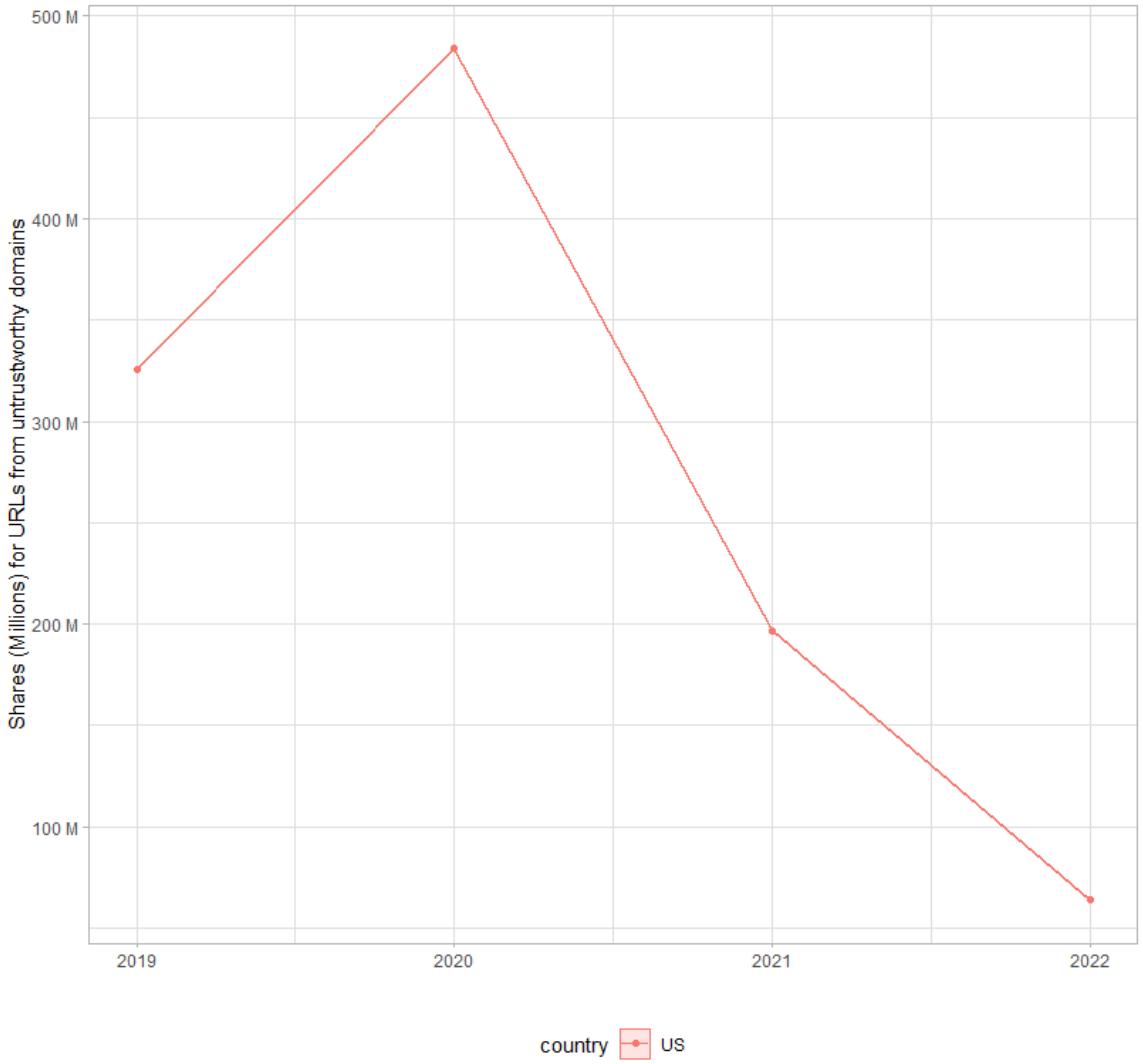

Figure 5. Number of *Shares* for URLs from untrustworthy sources

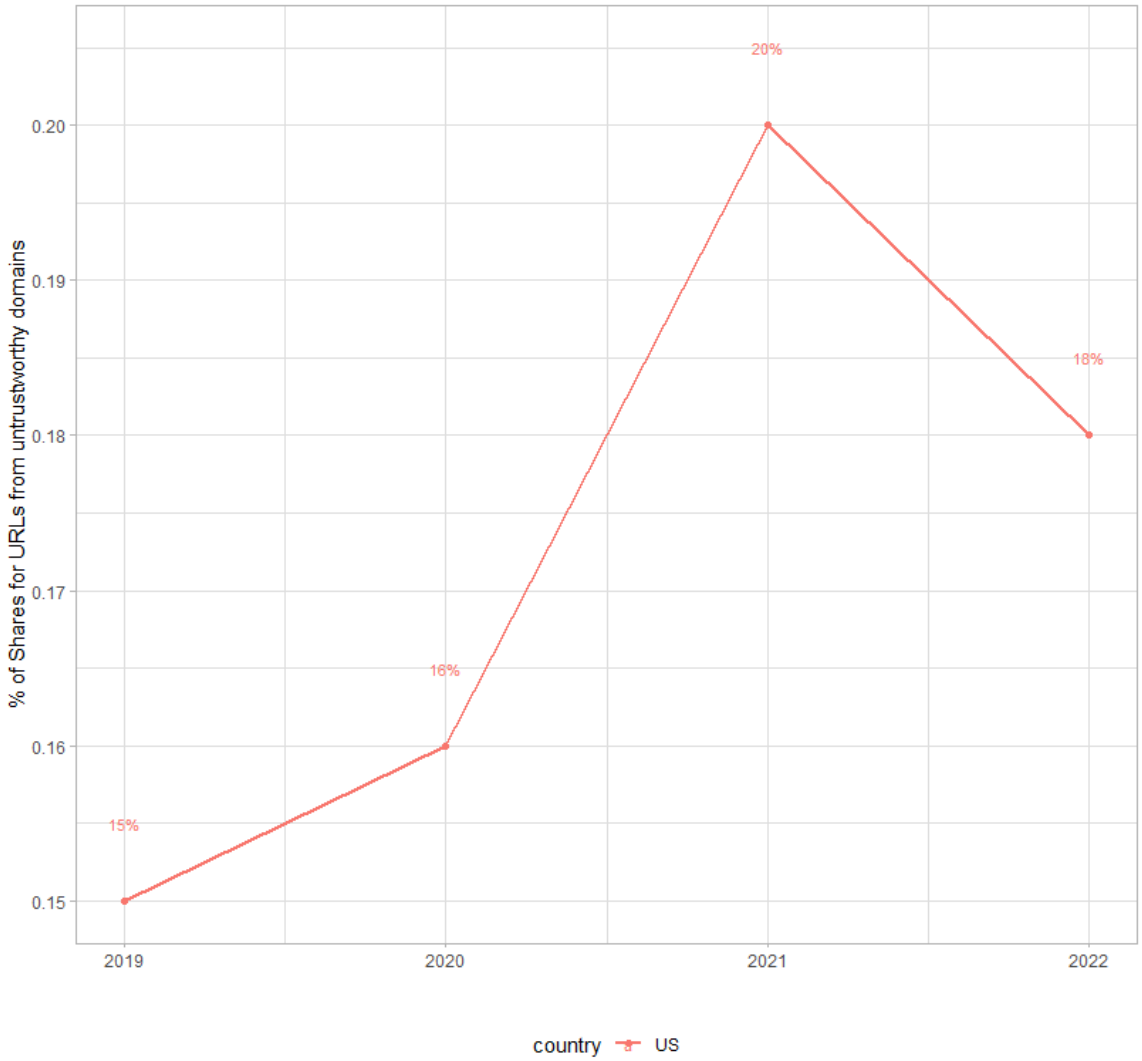

Figure6. Percentage of *Shares* for URLs from untrustworthy sources

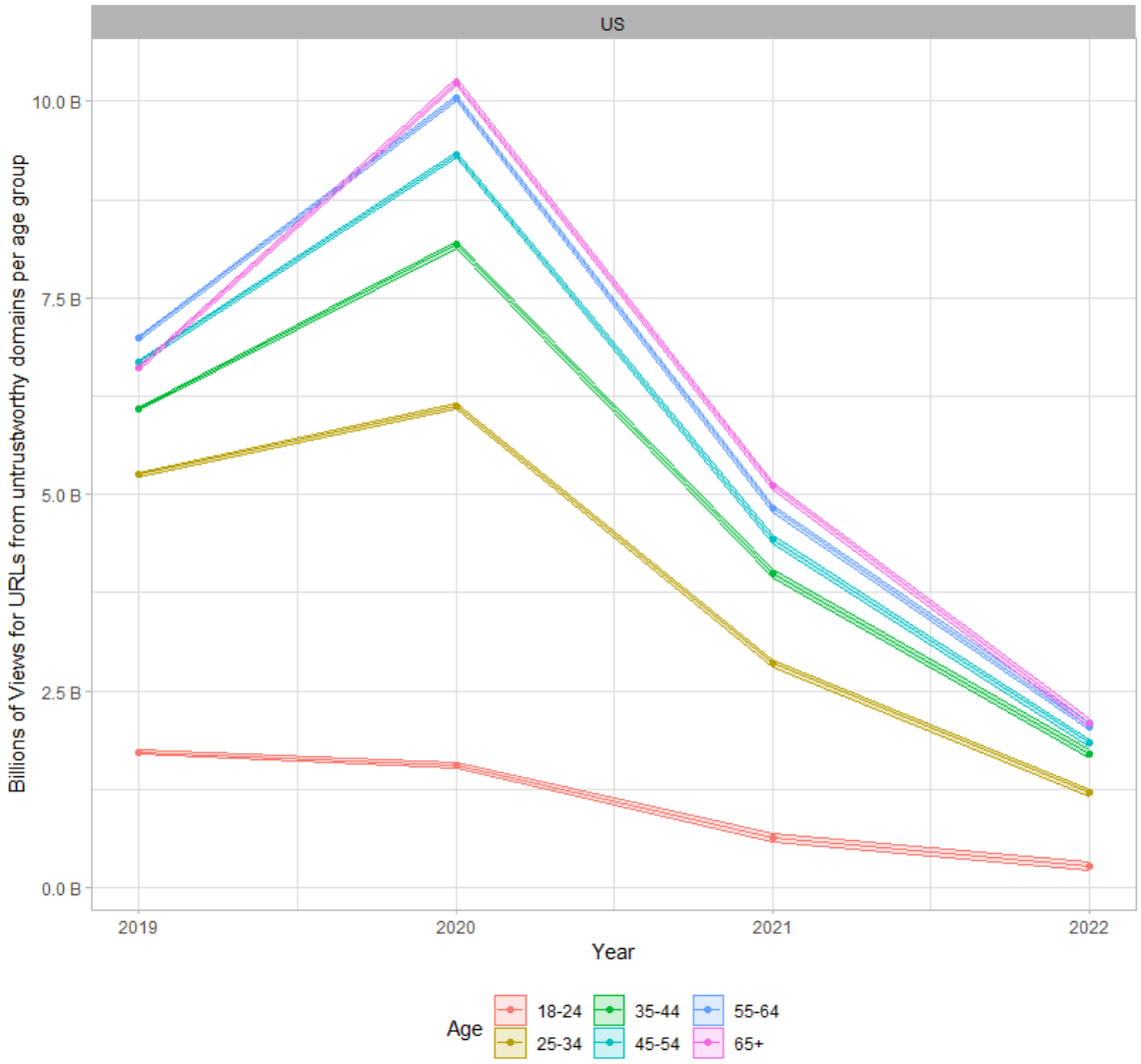

Figure 7. Number of *Views* per URLs from untrustworthy sources per age group

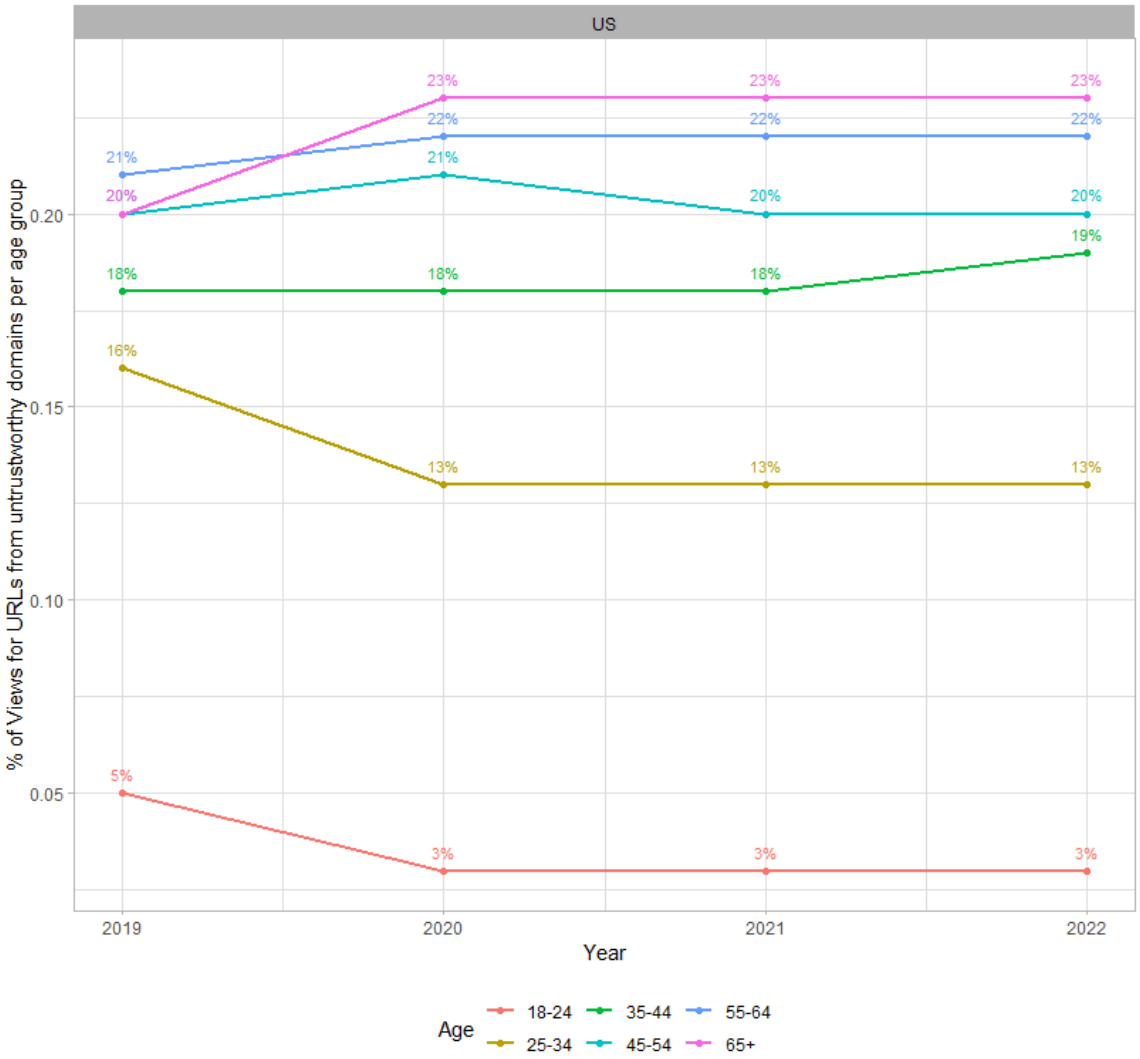

Figure 8. Percentage of *Views* per URLs from untrustworthy sources per age group

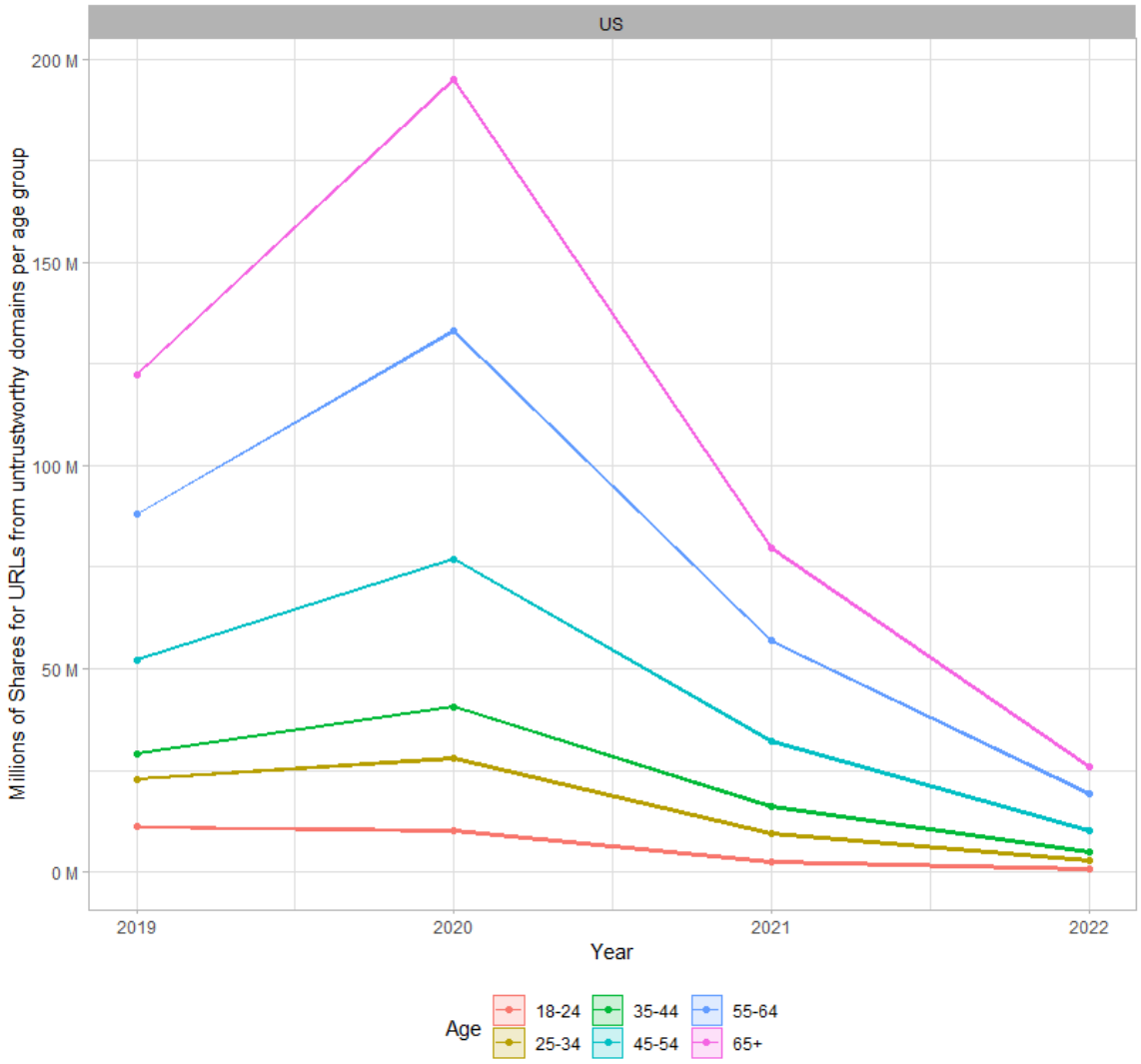

Figure 9. *Shares* of URLs from untrustworthy sources per age group.

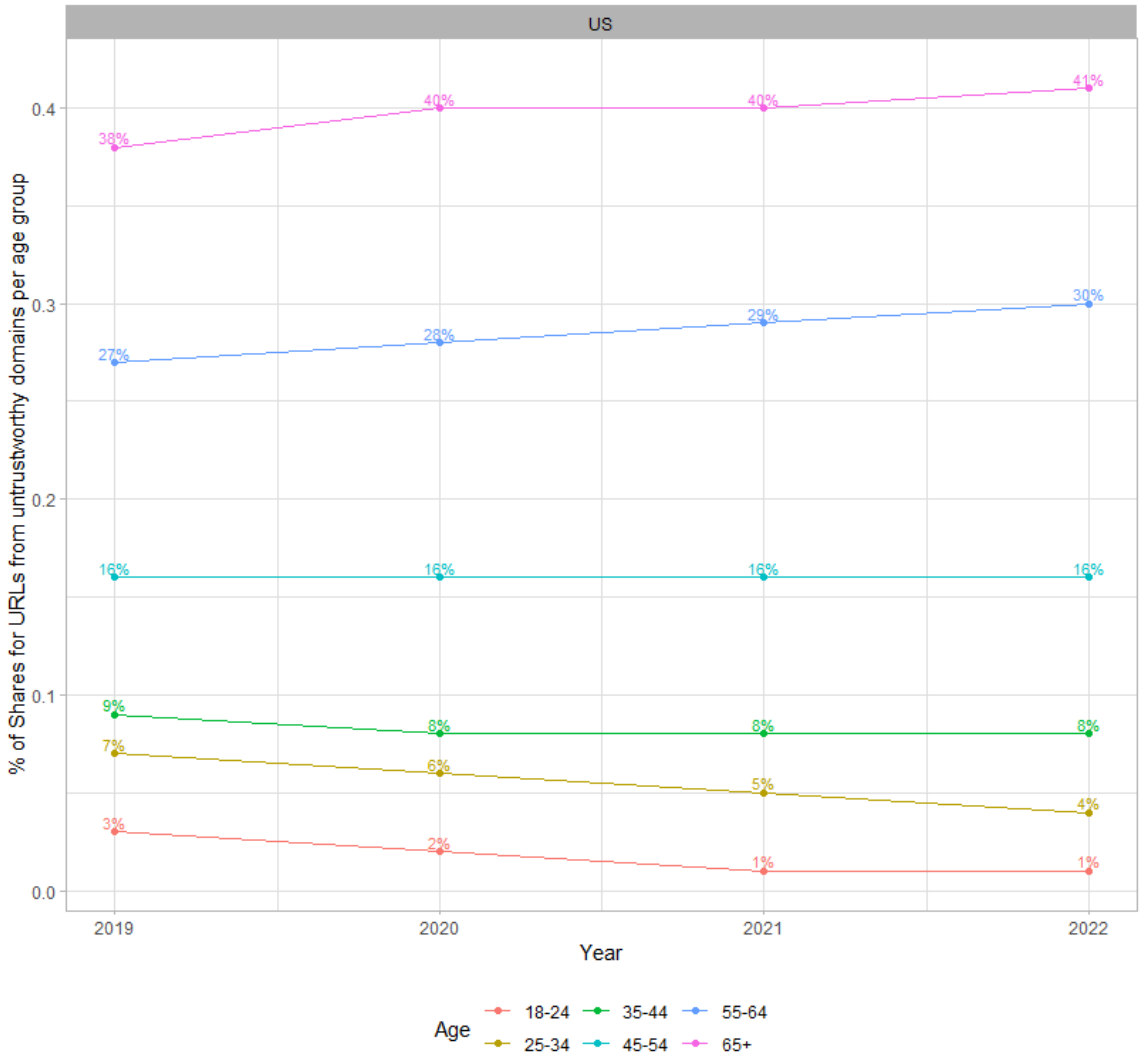

Figure 10. Percentage of *Views* per URLs from untrustworthy sources per age group

Appendix B

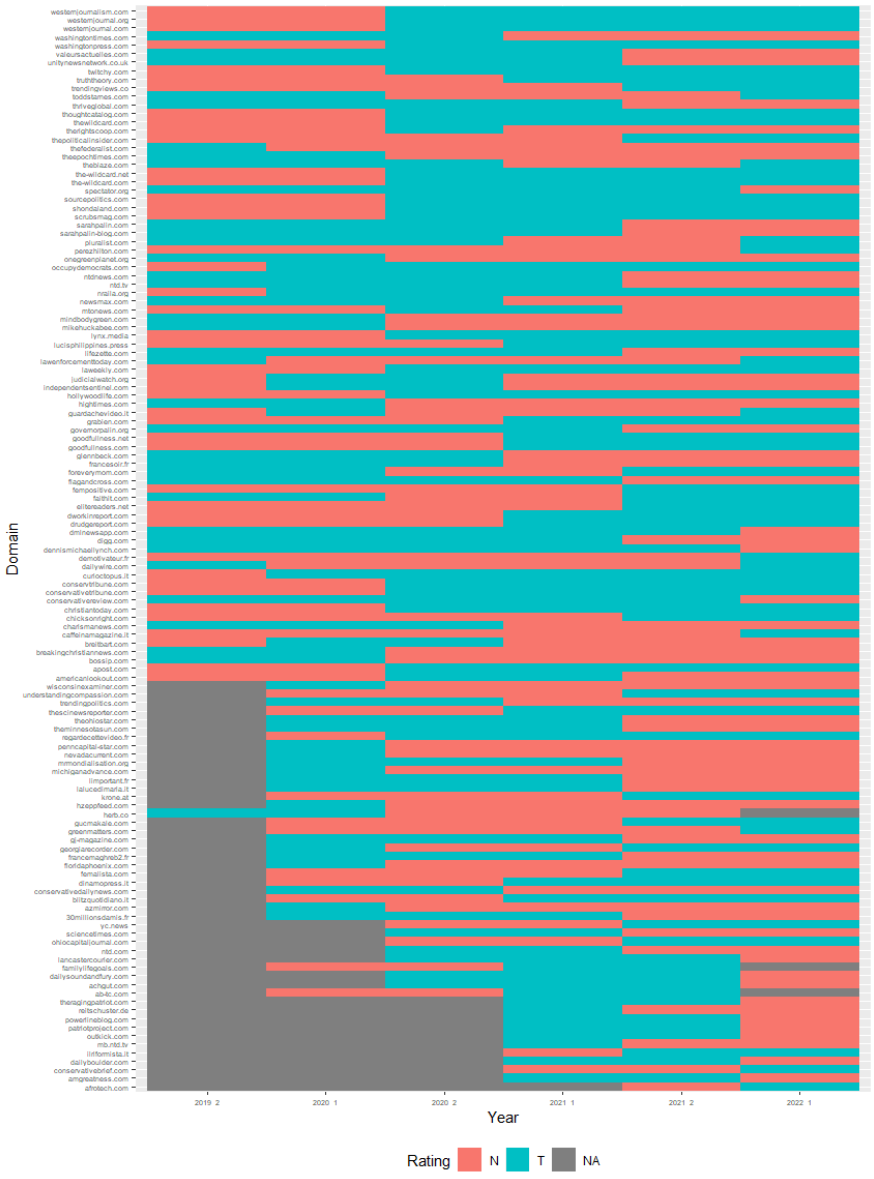

Figure 1. Newsguard domains, Rating over time