News sharing as a measure of media alignment

GÁBOR SIMONOVITS
ÁDÁM VIG
Central European University, Austria

In this note we introduce a new approach to measure media alignment derived from the story-sharing behavior of journalists. We use a large corpus of online news stories from two leading Hungarian news sites and estimate alignment scores for a large number of outlets that they cite. To the extent that journalists are more likely to cite ideologically proximate sources, our measure can be used to compare a large number of media outlets on a political — in our case government vs. independent — space. We demonstrate the use of this approach with two empirical applications. First, we show that our alignment scores successfully capture known ideological variation across outlets at a single point in time. Second, we demonstrate that quarterly estimates of alignment for a captured outlet change dramatically following an abrupt change in ownership.

Keywords: media alignment, news sharing behavior, media capture

Scholars and commentators have long worried about the fragmentation of the contemporary online media landscape and the possibility that it will lead to ideological echo chambers (Flaxman et al., 2016). At the same time, measuring the ideological slant of news outlets has proven elusive (Groeling, 2013) and current approaches come with important limitations. First, most content-based measures are costly because they necessitate the collection of news stories from each outlet to be coded by experts — or more recently crowdsourcing (Budak et al., 2016; Peterson et al., 2021; Allsides.com, 2018) and large scale text-analysis (Gentzkow and Shapiro, 2018).
To assess slant. Second, measures based on consumer choice — e.g. the likelihood that stories from a source are shared by liberals vs. conservatives \cite{Bakshy2015,Flaxman2016,Robertson2018} rely on proprietary data. As a result, most existing analyses rely on cross-sectional estimates of media content on a limited number of outlets.

In this note, we introduce a new measure of media content derived from the story-sharing behavior of journalists. Our approach rests on a simple insight: if journalists are more likely to cite and share stories from sources which lie closer to their own ideology \cite{Groseclose2005} then we can use such sharing behavior to infer the underlying alignment of the shared sources. Our approach builds on prior research by \cite{Adamic2005b}, who studied the ideological nature of story sharing in the American blogosphere as well as \cite{Groseclose2005}, who relied on the ideology of think-tanks cited by media outlets to estimate slant. Our study is also related to recent research by \cite{Kim2022}, who estimate media slant relying on the amount of screen time TV channels give to political actors as well as \cite{Barbera2015} who used follower-networks on Twitter to identify ideology of users. Closest to our study is \cite{Wihbey2019}, who connect the social networks of journalists with the media content they produce.

There are two key advantages of our approach. First, it relies only on the network of journalists and their sources — thus not necessitating the analysis of text. This makes our approach significantly more effective than approaches that rely on content analysis, even if they utilize a combination of crowdsourcing and machine learning \cite{Budak2016}. Second, as our application demonstrates, we can estimate alignment scores for a large number of outlets relying solely on news stories from two anchor outlets, assuming that journalists across those two outlets exhibit meaningful differences in their sharing behavior. Third, our approach is especially suitable to track the alignment of media outlets through time, facilitated by the intuitive comparability of our measure across time.

The context of our study is Hungary — a country often cited as an example of a new wave of democratic backsliding \cite{Serhan2020}. According to most observers, press freedom has steadily declined in the past decade \cite{Repucci2019} and a larger number of online media outlets have been captured by the government \cite{Szeidl2021}. Our empirical analysis relies on news stories published by two of the largest Hungarian online news sites. The first,
24, has been politically independent since its launch in 2000 and has emerged as one of the key independent online media outlets in Hungary. The second, origo, was acquired in 2016 by a firm linked to the cousin of the central bank’s governor (Szeidl and Szucs, 2021). Thus, in the time period — between 2016 and 2021 — the two outlets exemplify the content of the government and independent news ecosystems. Relying on the citation patterns of these two anchor outlets, we set out to characterize the alignment of other news sites cited in these two outlets.

**Data and Methods**

In order to study linking patterns, we first scraped all articles published on 24 and origo between 2016 and 2022. We extracted hyperlinks from each story and trimmed them so that we can identify individual websites the links were pointing to. We also have data on the exact date the article was published and the section it appeared on the news site. We coded the section of the news sites in which a story appeared and classified these sections to soft and hard news. Specifically, we classified news appearing in the domestic, international and economy sections as hard, while every other section (e.g. sports, tech, celebrities) as soft news. The resulting corpora included more than 650,000 published news stories from which almost 390,000 contained a hyperlink. We found that a large share of these citations are “self-citations” (they refer to another article published on the same news site) or refer to foreign sites (see Table A1).

We estimate alignment as the percentage difference between the shares an outlet received from origo vs. 24 (Bakshy et al., 2015). Specifically, denoting the count of citations a new site i received from origo and 24 as \( C_{iO} \) and \( C_{i24} \) respectively, we define our measure of alignment as

\[
A_i = 100 \times \frac{C_{iO} - C_{i24}}{C_{iO} + C_{i24}}.
\]

Note that calculating this measure does not require that we have any estimate of the ideological differences between the two anchor websites. For our measure to capture differences in alignment across the cited outlets, we simply need these two anchors to be sufficiently

---

1 All data necessary to reproduce our results are available [here](#).
different in their political outlook. In this sense, our approach is similar to Gentzkow and Shapiro (2010) who assess slant based on the similarity of a news outlet’s language to that of congressional Republicans or Democrats without an ex ante quantification of the ideological differences between these two groups of politicians.

We use two complementary strategies to validate our measure of alignment. First, we compare our measure of alignment to expert classification of 30 Hungarian news sites reported in Bátorfy (2020) which classifies outlets as either pro-government or independent, based on both their ownership and content. In our observed period, there were 45,408 articles on 24 and origo that contained a hyperlink pointing to one of the 30 news outlets (see Table A3 for a yearly breakdown). The advantage of this approach is that we can use off-the-shelf data on a large number of outlets compiled by respected experts. However, because the classification in Bátorfy (2020) is partially based on the subjective assessment of content, it remains unclear which dimension of alignment our measure captures.

As a second validation exercise, we focus on a single outlet, Magyar Nemzet (Hungarian Nation) which experienced a sudden change of ownership following the 2018 general election. Magyar Nemzet is one of the oldest Hungarian dailies which took an anti-government stance after a conflict between the government and its owner, oligarch Lajos Simicska, a college roommate of PM Orbán, former head of the Tax Authority (Szeidl and Szucs, 2021). However, after the landslide re-election of the government in April 2018, the daily was re-captured: Simicska sold all his interest in the Hungarian media market including Magyar Nemzet and later in February 2019 the pro-government daily Magyar Idők changed its name to Magyar Nemzet. Comparing the alignment of this single outlet before and after its capture helps us to explore if our measure can also pick up changes in media alignment resulting from capture.

It is important to note that our approach to measuring story-sharing is non-comprehensive. On the one hand, our reliance on hyperlinks means that we miss instances of story-sharing when links are not provided to the source. On the other hand, it is possible that sometimes journalists cite sources that they disagree with and may call the veracity of the source in question. However, note that our goal is not to provide an account of news-sharing, but instead to obtain a measure of alignment. Thus, we think that because providing hyperlinks to a cited outlet

\footnote{See Figure A2 for audience metrics.}
is a stronger signal of endorsement, focusing on these increases the validity of our estimates. Second, “negative” citations simply make the relative share of hyperlinks to an outlet from pro-government vs. independent sites a less reliable signal of alignment.

Results

We start our empirical analysis by summarizing alignment and citation counts across our sample. Figure 1 reveals important variation across the outlets cited in origo and 24. Some large independent news sites such as index, hvg and nepszava received a very large number of citations (in the magnitude of thousands), almost exclusively from 24. Equally highly cited are online tabloids such as bors and bikk which are not aligned with either side. Government-backed political news sites such as 888 or pestriscrak are completely aligned with origo, but receive fewer citations. Finally, some government-funded sites focusing on the economy (VG and figyelo) are relatively moderate in their alignment. The y-axis gives an idea about order magnitude of citations different news sites received in the two anchor outlets.

Figure A1 shows the empirical CDF of the total number of citations received by domains. For example, there are 175 sites that received 100 or more citations from 24 and origo, but only 47 that received 500 or more.
Figure 1. Alignment scores for Hungarian news sites

*Note.* Alignment for sites with total citation over 30. Period: 2016.01.01 - 2021.12.31

Figure 2 compares our measure of alignment across outlets classified by experts as pro-government vs. independent. For the sake of simplicity, we compute alignment scores based on news stories published on *origo* and *24* between 2019 Q2–2020 Q2, the most recent full year in our sample when no source experienced ownership change. Figure 2 shows a huge degree of polarization in terms of linking behavior. First, stories by pro-government media outlets were much more likely to be shared by *origo* compared to 24. Second, independent outlets were almost never quoted by *origo.*
This analysis confirms that our measure of media alignment computed by citation patterns in two anchor outlets can distinguish between pro-government and independent sources. Moreover, our results also demonstrate the usefulness of our measure compared to the binary classification. Our more fine-grained measure is able to identify variation in alignment within pro-government outlets. For instance, VG and 888 are both owned by KESMA, a government-funded media conglomerate. While the former is a relatively moderate outlet covering business news, the latter focuses mostly on opinion pieces attacking the opposition. In comparison, our measure reveals no such variation across independent outlets. This implies that — at least from the point of view of government controlled media — independent outlets are relatively homogeneous.
Figure 3. Quarterly estimates of alignment score of *Magyar Nemzet*

Note. *Magyar Nemzet* was not operating between April 2018 and February 2019, therefore this period is left out from the analysis. 95% confidence intervals are calculated from the standard error of the mean.

Figure 3 plots the proportion of links pointing from *origo* and *24* to *Magyar Nemzet* before and after its re-capture. Strikingly, before the capture of the outlet — and especially in the period leading up to the 2018 election — the pro-government *origo* virtually ignored stories appearing on *Magyar Nemzet*, compared to *24*. However, immediately after the transition, its relative coverage by the two outlets flipped: by the end of 2019 stories from MN were much more often cited in *origo*. All in all, this analysis again shows that the citation links alone were able to recover a quite dramatic change in alignment and provide a proof-of-concept that this measurement approach could be used as a forensic tool to detect capture.

**Conclusion**

In this note, we introduced a new measure of media alignment derived from the story-sharing behavior of journalists. Under the assumption that the choices journalists make between dif-
Different sources reflect ideological motives, our measure can be used to place a large number of media outlets on an ideological continuum with minimal data and modeling assumptions. The only identifying assumption we make is that our anchor news sites are selected from opposing sides of a highly polarized media environment. Studying the political alignment of media outlets in a setting with a rapidly deteriorating degree of media freedom and ever-expanding state involvement in communication, our approach can aid the assessment of how a rapidly-changing ownership structure shapes content.

At the same time, our focus is likely to limit the generalizability of our approach in settings where it is harder to identify a principal axis of political differentiation between media outlets or in settings where the use of hyperlinks is less prevalent (e.g., traditional print media or TV). That being said, the quite minimal requirements of our approach in terms of data (news stories and hyperlinks from two ex ante different anchor outlets) makes our approach a viable candidate in settings where previous research identified a single dimension of political competition — as it is usually assumed in two-party systems such as the U.S. or settings such as the one studied here, featuring a dominant party. We believe that our validation exercises show that our approach can complement existing measures of media bias and should be put to test in other settings.

Acknowledgments

We thank the reviewers and editors of the journal for their helpful comments.

References


Appendix

Table A1: Number of articles and citations in our sample

<table>
<thead>
<tr>
<th></th>
<th>24</th>
<th>origo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of articles</td>
<td>377839</td>
<td>279001</td>
<td>656840</td>
</tr>
<tr>
<td>Article contains citation</td>
<td>265229</td>
<td>131191</td>
<td>396420</td>
</tr>
<tr>
<td>'Self citation'</td>
<td>66011</td>
<td>34452</td>
<td>100463</td>
</tr>
<tr>
<td>Foreign outlets</td>
<td>65479</td>
<td>26878</td>
<td>92357</td>
</tr>
<tr>
<td>Hungarian media sample</td>
<td>34232</td>
<td>11176</td>
<td>45408</td>
</tr>
</tbody>
</table>

Table A2: Total number of outbound links from each site by year

<table>
<thead>
<tr>
<th></th>
<th>24</th>
<th>origo</th>
</tr>
</thead>
<tbody>
<tr>
<td>year</td>
<td>2016</td>
<td>54429</td>
</tr>
<tr>
<td></td>
<td>2017</td>
<td>54933</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>48833</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>39074</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>34746</td>
</tr>
<tr>
<td></td>
<td>2021</td>
<td>27179</td>
</tr>
</tbody>
</table>
Table A3: Number of outbound links from each site by year to our Hungarian media sample

<table>
<thead>
<tr>
<th>year</th>
<th>page 24</th>
<th>origo</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>7657</td>
<td>1169</td>
</tr>
<tr>
<td>2017</td>
<td>6848</td>
<td>1261</td>
</tr>
<tr>
<td>2018</td>
<td>5887</td>
<td>2183</td>
</tr>
<tr>
<td>2019</td>
<td>5707</td>
<td>2533</td>
</tr>
<tr>
<td>2020</td>
<td>4260</td>
<td>2453</td>
</tr>
<tr>
<td>2021</td>
<td>3873</td>
<td>1577</td>
</tr>
</tbody>
</table>

Figure A1. Empirical cumulative distribution function of total citations per domain
Figure A2. Audience of Hungarian online media sites

*Note.* We calculated the share of daily internet users as a ratio of daily average users of a news site over the total number of real users in Hungary. Source: gemiusAudience (2022)

Figure A3. Alignment scores for Hungarian online news sites by content type

*Note.* 95% confidence intervals are calculated from the standard error of the mean. Period: 2019.07.01–2020.06.31
Figure A4. Quarterly estimates of alignment score of *Magyar Nemzet* by content type

*Note.* *Magyar Nemzet* was not operating between April 2018 and February 2019, therefore this period is left out from the analysis. 95% confidence intervals are calculated from the standard error of the mean.