Article

Writer movements between news outlets reflect political polarization in media

Nick Hagar [iD]
Northwestern University, USA

Johannes Wachs
Vienna University of Economics and Business, Austria; Complexity Science Hub Vienna, Austria

Emőke-Ágnes Horvát
Northwestern University, USA

Abstract
Digital news outlets rely on a variety of outside contributors, from freelance journalists, to political commentators, to executives and politicians. These external dependencies create a network among news outlets, traced along the contributors they share. Using connections between outlets, we demonstrate how contributors' publishing trajectories tend to align with outlet political leanings. We also show how polarized clustering of outlets translates to differences in the topics of news covered and the style and tone of articles published. In addition, we demonstrate how contributors who cross partisan divides tend to focus on less explicitly political topics. This work addresses an important gap in the media polarization literature, by highlighting how structural factors on the production side of news media create an ecosystem shaped by political leanings, independent of the priorities of any one person or organization.

Corresponding authors:
Nick Hagar, Communication Studies, Northwestern University, Evanston, IL 60208-0001, USA.
Email: nhagar@u.northwestern.edu

Emőke-Ágnes Horvát, Communication Studies, Northwestern University, Evanston, IL 60208-0001, USA.
Email: a-horvat@northwestern.edu

Introduction

Political polarization in digital news consumption is well established. While the exist-ence of self-reinforcing isolation that is strong enough to *override* individual agency—so-called "filter bubbles"—is hotly contested, there is substantial evidence of patterns of media consumption aligned with political identity in the United States (Bakshy et al., 2015; Macy et al., 2019; Shi et al., 2017). Whether by an explicit recognition of political identity or the implicit association of certain behaviors and traits with a politically aligned community, in-group identity drives members of opposing political groups into silos (Coleman, 1988; Dvir-Gvirsman, 2017). However, this work only explores con-sumption dynamics.

As a result, researchers have often sought to test the efficacy of cross-cutting content and conversation in reducing polarization, which exposes members of opposing groups to challenging viewpoints (Heatherly et al., 2017). The idea of intentional exposure to conflicting views finds its way into proposed interventions. Past work on "bursting" fil-ter bubbles introduced tools that allow users to monitor the political polarity of their news consumption, for example, or to seek out the opposing views on important issues (Flaxman et al., 2016; Resnick et al., 2013). In an effort to support consumers, other work imagines alternative modes of suggesting media to users through recommendation systems (Helberger, 2019). These interventions address exposure to news, but not its production.

In news production, research often focuses on the individual priorities and profes-sional norms countervailing to polarization. In particular, researchers pay close attention to the professional norm of objectivity within journalism and how it gets enacted to mini-mize partisan influence (McChesney, 2003; Schudson and Anderson, 2009). This per-spective focuses on the individual journalist as the counterpoint to polarization in news media, as they carefully construct coverage with an eye toward fairness, balance, and accuracy (Ryan, 2001). And while this view acknowledges the role of news producers in polarized environments, it also fails to address the structural forces shaping that production.

Research that does take a structural approach finds ample evidence of polarization in the distribution of digital news. The earliest political blogs demonstrated an overwhelm-ing preference for linking to and engaging in dialogue with like-minded outlets (Adamic and Glance, 2005). Similarly, work on the dissemination of digital election coverage found an isolated, self-reinforcing group of right-leaning outlets, one that largely disen-gaged itself from the rest of the news ecosystem (Benkler et al., 2018). Third-party social media platforms used for news distribution seem to largely amplify these producer pref-erences (Bakshy et al., 2015; Benkler et al., 2018; Wihbey et al., 2019). Therefore, even in a world where media consumers have the agency to cross political boundaries and discover opposing views, the distribution of that media operates in such a way that dis-courages cross-cutting exposure. Polarization is an integrated problem.

This study attempts to extend that structural view into the production of digital news itself, by examining the movement patterns of digital news contributors from outlet to outlet. We treat these movements as a network, in which edges form between outlets who share contributors. Using thousands of online news stories from 13 digital outlets—with

manually validated information about the journalists, freelance writers, and political actors who wrote them—we demonstrate clear patterns in the trajectories contributors take when they write for multiple outlets. We then show that contributor trajectories across outlets align with the underlying political leanings of the outlets in our sample, more than any other characteristic of the outlets or their audiences. Furthermore, we show that those politically aligned clusters differ in internal structure, with a highly interconnected cluster of right-leaning outlets. Finally, we link the political clustering present in this network to the content contributors produce. We find that, not only do both the topics writers cover in each cluster and the ways in which they present this content differ, but also that contributors that move between clusters tend to be less explicitly political. Taken together, these findings challenge the value of bridging content in reducing the polarization of digital media consumers. Rather, they demonstrate the need for a more nuanced examination of media *producers* and their underlying motivations.

Background

In news' shift to digital production, newsrooms increasingly rely on freelancers and other outside contributors. This reliance on freelance work creates an open and competitive market for news production. Individual news contributors in this market must develop strategies for successfully navigating news outlets, especially in the midst of extreme political polarization. Those strategies can appear in the movement patterns of contributors from outlet to outlet, as well as the topics they cover. We posit that such contributors who successfully cross political lines offer potential strategies for reducing polarization.

Digital journalism's shift to outside contributors

For digital news outlets, the shift to outside contributors is largely the result of resource constraints—layoffs push news producers into less secure work positions, such as freelancing, as news revenue continues to decline (Cohen et al., 2019). For professional journalists, freelancing has become the major form of news production work in some places (Hayes and Silke, 2018). Because freelancing relies heavily on individuals pitching news pieces to editors, this shift also represents an increased responsibility for contributors to decide what topics to cover, and for which news outlets (Rosenkranz, 2018).

At the same time, the type and focus of digital news coverage have moved away from traditional event-centered reporting. Rather than present an isolated account of a newsworthy occurrence, reporting often emphasizes analysis and interpretation of events in a broader context (Barnhurst and Mutz, 1997). That change has only been accelerated by blogging and news production on social media, opening up news analysis to a wide array of contributors outside the traditional journalism field (Lowrey, 2006; Singer, 2005). These concurrent shifts create a less formalized sphere of news production, one that extends beyond the traditional professional class of journalists. In turn, news contributors expand to include not only journalists, but also writers, academics, politicians, and other individuals who generate news analysis and interpretation for online audiences.

Such an open, digital environment for news production has two key implications for contributors. First, it creates rapid and constant competition, in which all news producers

have equal access to technology and the ability to produce content (Munger, 2020). Second, external modes of labor like freelancing are most often temporary, meaning workers must constantly move from organization to organization (Berton et al., 2011). In these conditions, news producers must develop and carry out successful strategies for shifting from one position to another, in order to remain competitive.

For full-time news contributors, freelance work is often synonymous with precarious work. Income becomes far less stable, and freelancers must bear additional costs associated with reporting (Gollmitzer, 2014; Salamon, 2020). These challenges add additional pressure to finding successful strategies, since freelancers' livelihoods depend on a constant flow of new work. In particular, success in freelance work depends on the *patterns* of that work—what opportunities freelancers take, and which others arise in conjunction (Leung, 2014). In news production, those patterns appear most prominently through in which outlets contributors publish articles. Given that emphasis, we ask,

> *RQ1*. What patterns do news contributors, particularly external contributors, follow when moving among outlets?

Polarized coverage for fragmented audiences

Digital news exists in a fragmented ecosystem. In contrast to mass media like broadcast or print, news consumers split their attention among a wide variety of sources (Webster, 2016). Audiences splinter into small groups as low-barrier, low-cost news sources tailor to their specific interests and identities (Taneja and Wu, 2018).

One prominent dimension of this fragmentation is political polarization. Past work has argued for the existence of "filter bubbles" or "echo chambers," completely isolated spheres of media consumption. The evidence for these kinds of all-encompassing divisions in news consumption is limited (Bruns, 2019). However, political beliefs and preferences are a clear *factor* for consumers in news source and story selection. People generally prefer to read news that align with their political positions (Flaxman et al., 2016). Selective exposure along partisan lines also shapes readers' perceptions of news source credibility, thereby impacting what news they decide to engage with (Tsfati et al., 2014).

In turn, audience preferences feed into increasingly polarized news outlets. Digital news media are heterogenous, with varying agendas, topical focuses, and speeds of reporting (Harder et al., 2017). Those differences serve to distinguish outlets from each other, drawing in different audiences depending on their preferences. As a consequence, news outlets can become more polarized in response to competition, as a strategy for capturing market share (Mitchelstein and Boczkowski, 2009; Mullainathan and Shleifer, 2005). Even in markets with news consumers who do not explicitly seek out polarized viewpoints, models show that newspapers are served well by taking clear political stances (Anderson and McLaren, 2012).

At the same time, many news organizations and individual journalists go to great lengths to avoid the appearance of any partisan position or influence. As journalists developed a cohesive professional identity throughout the 20th century, industry-wide

norms developed that shifted news' presentation (Schudson, 2001). In particular, journalists developed a strong link between professionalism and objectivity (Schudson and Anderson, 2009). This objectivity seeks to balance the perspectives and positions represented in news coverage, rather than advancing any particular narrative (Ryan, 2001). It is important to note that objectivity is often an ethical ideal, not a value reflected in published news. Even so, objectivity (or at least its appearance) is a valuable strategy for news organizations, as it enhances the credibility with readers that they rely on for sustainable growth (Singer, 1997).

Polarization and objectivity set up conflicting accounts of how news organizations position themselves in competitive markets, and they have distinct implications for the strategies outside contributors might pursue. Coverage polarization incentivizes politically aligned niches of topics and perspectives, while objectivity encourages broadly salient, more neutral information. These contrasting potential strategies lead to our second research question:

RQ2. Do the topics addressed by contributors vary, depending on an outlet's political leaning?

Structural polarization and crossing political boundaries

Structural factors are a major driver of polarization in the production, distribution, and consumption of digital news. As mentioned above, market competition pushes news outlets gradually further away from the center of the political spectrum, in an effort to better capture certain subsets of news readers (Anderson and McLaren, 2012; Munger, 2020). Algorithmic intermediaries also reinforce polarization through distribution, in that they reward more extreme offerings through recommendations (Blex and Yasseri, 2020; Ribeiro et al., 2020).

Even beyond these mechanisms, researchers have identified *network-based* characteristics of the digital news ecosystem that reinforce polarization. In particular, Benkler et al. (2018) examine the pathways along which news stories travel among outlets. They find that, across several modes of distribution, right-leaning outlets tend to amplify each other's stories most often. They are also isolated from the rest of the media ecosystem, lacking ties, in terms of hyperlinks and social media sharing, to center- or left-leaning outlets. Outlets also diverge in the topics they cover, depending on their political leaning. In this way, Benkler et al. (2018) paint a picture of a media ecosystem that is polarized, not because of any individual agent or behavior, but because of a broad structure of distribution. This account fundamentally addresses the *portability* of news stories and perspectives among audiences.

Outside news contributors add an important piece to this structural account, offering potential novel approaches for combating polarization. There has been much focus on "bursting" news readers' filter bubbles, by giving them indications of the partisan slant of their consumption habits or intentionally exposing them to opposing viewpoints (Flaxman et al., 2016; Resnick et al., 2013). This approach attempts to tackle partisan isolation in the digital news ecosystem by focusing on cross-cutting exposure in

consumption. News *producers* can provide a productive complement to that approach in a couple ways. First, from a networks perspective, movement across groups is key for valuable information flow (Granovetter, 1973). Cross-pollination increases exposure to novel ideas and productivity (Granovetter, 1973; Montgomery and Nyhan, 2017; Vedres and Stark, 2010). From an individual's perspective, bridging the gap among isolated groups presents an opportunity for potentially lucrative professional connections (Burt, 1992). By crossing political lines, news contributors may be rewarded with new connections, perspectives, and editorial approaches. They, therefore, have incentives to cross political lines that consumers do not.

Second, cross-cutting perspectives sometimes create unintended consequences, causing the recipient to become less receptive to opposing views (Bail et al., 2018). To some extent, that negative outcome may stem from the producer's choice of topic or perspective. In environmental news, some frames (e.g. focusing on public health) are more effective than others at engaging climate change skeptics (Bolsen and Shapiro, 2018). More broadly, news consumers' choice of topics often stems from social divisions like political affiliation (Knobloch-Westerwick and Meng, 2009; Tewksbury and Riles, 2015). These findings suggest that cross-cutting news must be tailored to its recipients, such as by more strategically selecting topics. By examining contributors who have successfully produced news across polarized audiences, we may get a better sense of which news topics appeal across political lines. To examine this possibility, we ask,

> *RQ3*. Do contributors who cross political boundaries cover different topics from those who do not?

Data

To identify news contributors who moved from outlet to outlet over time, we relied on a manual review of cleaned byline data from a broad sample of news articles. We started with a publicly available data set of news articles scraped from the homepages of major publishers (Thompson, 2017). The full data set contains 131,860 articles, published between June 2014 and July 2017, across 14 outlets: the New York Times, Breitbart, CNN, Business Insider, the Atlantic, Fox News, Talking Points Memo, BuzzFeed News, National Review, New York Post, the Guardian, NPR, Vox, and the Washington Post.[1] These outlets provide wide variation along a number of key characteristics; not just political valence, but also audience size, institutional prominence, geography, and the nondigital publishing channels they utilize. While these outlets represent only a subset of all digital news media, relying on a pre-made set of news outlets precludes any bias that might arise from our own selection criteria. The results we show below are also not dependent on this particular set of outlets, as removing outlets at random does not change the outcome of our analysis (see Supplemental material, Section 1.2 for details of this robustness check).

We filtered this full data set in three ways. First, we removed any records with author fields that do not correspond to a person's name. To accomplish this, we removed any stories with bylines that matched generic phrases (e.g. "the editors," "anonymous"), outlets (e.g. "The Associated Press," "NPR staff"), or that contained bylines with multiple
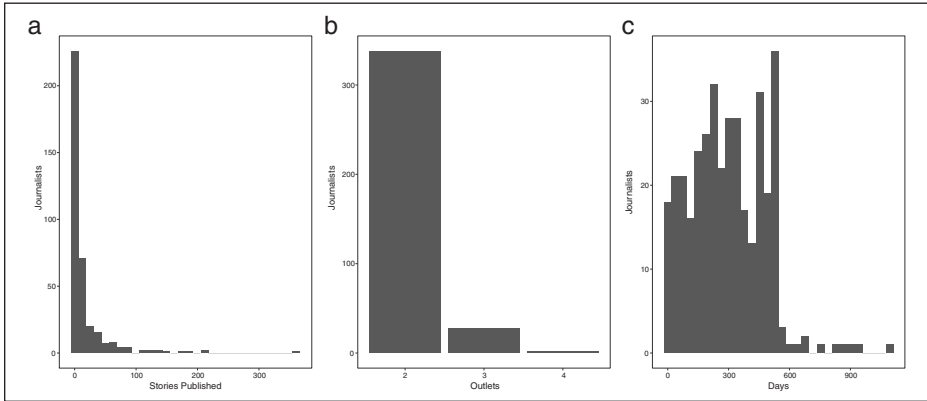
Figure 1. Distributions of (a) story counts per contributor, (b) unique news outlets that published each contributor, and (c) the number of days from each contributor's first to last story in our data.

names. Second, because we are interested in movement between outlets, we limited our sample to articles by contributors who appeared in at least two outlets in the data set. Because Business Insider does not share any contributors with other outlets, we removed it from our analysis. Finally, we manually verified the remaining bylines, disambiguating cases of multiple people sharing the same name. This process identifies 368 contributors, who wrote 6032 articles and are the focus of our analyses. While these articles represent only 4.6% of our initial set, they are consequential in that they represent all inter-outlet movement within our sample. We are able to capture thousands of cases in which contributors move between outlets, laying the groundwork for a robust network model.

We next manually coded each contributor depending on their professional background. First, there are three types of professional journalists in our sample: Freelancers who write for multiple news outlets (97 names), journalists who are full-time employees at an outlet but write for at least one other (67 names), and journalists who move from one full-time job to another over the course of our sample (43 names). Overall, then, 207 of the contributors in our sample are professional journalists of some kind. We also find a couple groups who write pieces in support of particular issues or causes. This includes activists and think tank members (62 names) as well as political commentators and politicians (27 names). Finally, there are smaller groups of academics (33 names) and authors (33 names) who write pieces primarily to promote their work. An additional six contributors do not fit into these categories.

Within our sample of 368 identified contributors, we analyzed publishing patterns for each individual (Figure 2(a)). Figure 1 presents the volume of stories, number of outlets published in, and number of days from first to last story for each contributor. We find that many contributors write across a couple outlets, and the majority publish fewer than 100 stories. However, there is widespread variation in how long contributors are actively publishing stories throughout our sample, indicating differences in their tempo of activity.

Table 1. Measures used for classifying outlets, along with their sources.

Measure	Source
Total unique users	ComScore
Total time read	ComScore
Median reader income	ComScore
Median reader age	ComScore
Audience household size	ComScore
Audience children (yes/no ratio)	ComScore
Audience race	ComScore
Audience gender (M/F ratio)	ComScore
Audience region	ComScore
Outlet medium	Manual coding
Political leanings	See Bakshy et al., 2015

At the outlet level, we use a variety of characteristics to evaluate our eventual network's structure (Table 1). These characteristics fall under three broad types of measures, and they broadly capture information about the outlet and the size and composition of its audience. First, we used two measures of outlet size—total unique users and total time users spent reading each outlet—as these are often utilized in studies attempting to classify news outlets (e.g. Hindman, 2018; McCombs and Winter, 1981). Second, we used a variety of audience characteristics—household size, whether or not a household has children, the race and gender composition of the audience, where the audience is located, and the median reader age and income. All measures except income report audience size, in terms of unique users, for certain groups. For example, household size reports total users for households with 1–2 and 3–5 members. All measures except region, age, and income report two groups, so we calculated the ratio of one group to the other to reduce the number of classifications. For audience region, we retained the largest region by number of users. Audiences often identify with the news they read, particularly in alternative media, which could create observable clusters of outlets within shared communities (Benkler et al., 2018; Chiricos et al., 1997; Couldry and Curran, 2003; Kaiser et al., 2019). Third, we utilized a couple outlet characteristics: traditional publishing medium and political leaning. Publishing medium is a common classification approach for comparative work (e.g. Boczkowski and de Santos, 2007; Maier, 2010), and while the articles we analyze are all published online, outlets still maintain other dominant distribution media that may inform their overall publishing strategy. We utilized the outlet political leanings generated by Bakshy et al. (2015) in examining polarization on Facebook.

ComScore provides all our audience measures. For these, we obtained the mean monthly value of each measure from January 2016 to December 2017, when most articles were collected. We then divided each continuous measure into three groups, each of the size $\frac{(max - min)}{3}$. For political leaning, we use a measure devised by Bakshy et al. (2015). They analyze the self-reported political affiliation of Facebook users, assigning news articles an average alignment score based on who shared each article. They then
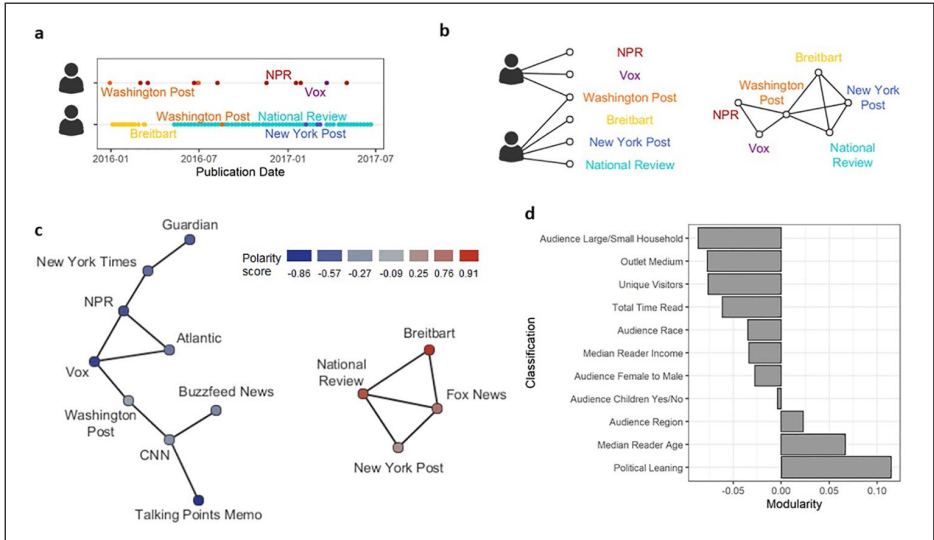
Figure 2. Mapping contributors' trajectories between news outlets. (a) Example trajectories indicating the outlets for which the selected two contributors wrote articles. (b) Illustration of how we build the network of outlets based on which pairs of outlets shared contributors. (c) Significant connections between news outlets ($Z > 1.96$), colored according to political leaning scores assigned by Bakshy et al. (2015). (d) Modularity for each outlet classification applied to the outlet network.

average those article-level scores at the website level, producing a site-level alignment score ranging from −1 (left-leaning sharers) to 1 (right-leaning sharers). Bakshy et al. (2015) coarsen this measure by quintiles. We follow a similar approach, treating sites that fall outside of their central quintile as left- or right-leaning.

Methods

Here, we describe the analytic approach we used to examine the network between outlets that maps news contributor movement, as well as to investigate the topic and tone of stories published.

Contributor network structure

To evaluate the network structure of contributor movement, we constructed a bipartite network of writers and outlets, utilizing only our filtered set of writers. A writer and an outlet shared an edge if at least one of the writer's stories appeared in the outlet within our sample. From this network, we constructed a one-mode projection for outlets, in which outlets share an edge if at least one writer published an article in both during the considered time frame (Figure 2(b)). Edges were weighted according to the number of writers outlets share. The number of articles per contributor across outlets also varies;

however, the article count is much less relevant to our research question than the number of contributors shared. In total, 31 pairs of outlets did not share any contributors.

We then compared this weighted projection to randomized bipartite networks of news outlets and contributors, to check for significant connections between outlets. We used a standard Markov Chain Monte Carlo (MCMC) algorithm to build random networks from the observed bipartite network (Cobb and Chen, 2003; Newman, 2001; Rao et al., 1996; Tumminello et al., 2011; Zweig and Kaufmann, 2011). Because our network is bipartite and has a skewed degree distribution, the MCMC approach provides the most exact null model comparison (Schlauch et al., 2015). We attempted $m \times log_e(m)$ edge switches, where m is the number of edges in the bipartite network (i.e. 784) to obtain a single instance of a random network that has a substantially different structure than the observed network (Gionis et al., 2007). We generated a set of 10,000 such random networks. Our Supplemental material (Table 1) shows the significance of edges in the projection; we retain edges where $Z > 1.96$. This procedure generates distinct and disconnected clusters of outlets.

Classification evaluation

To evaluate the uncovered network clustering, we compared it to classifications from the mixture of audience and outlet characteristics described in the "Data" section (Table 1). We treated each of these classifications as though they were partitions within our observed network, then used that partition to calculate the modularity of the classification. Ranging from −1 to 1, modularity describes the extent to which a network is split into distinct clusters (Newman, 2006, 2010). By mapping classifications onto our observed network as though they were clusters, essentially, we calculated the extent to which each one aligned with the observed structure.

Topic modeling

To analyze how the network structure of the journalism marketplace relates to content, we employed topic models (Mohr and Bogdanov, 2013). Specifically, we fit a topic model to the articles in our data set with Latent Dirichlet Allocation (LDA; Blei et al., 2002) using the Mallet program (and its default parameter settings) (McCallum, 2002), accessed through the gensim library (Řehůřek and Sojka, 2010) of the Python programming language. The result assigned each document a vector of topics, allowing that a document consists of a combination of topics and that words appear in multiple topics. The number of topics is a parameter of the LDA algorithm—in the article, we presented results from a model set to find 20 topics. We found similar results, namely significant differences in topic prevalence across the clusters of outlets when we fit the model to 15, 25, and 30 topics.

Before we fit the topic model, we processed the text of the articles. We extracted each word or token, lower-casing all characters and removing apostrophes. We removed English-language stop words and stemmed the remaining words using the Porter stemmer. We removed words that occurred in more than one-third of articles and those that occurred in less than 10 articles in our corpus. The general goal of these steps was to

reduce the noise of the data while keeping its main signals relevant to our research intact (Hopkins and King, 2010). The order of our processing pipeline is in line with recent recommendations of best practices in topic modeling for communication research (Maier et al., 2018).

Our network analysis finds a clear two-cluster structure, one that maps onto outlets' political leanings. We checked whether there are significant differences in content between our derived outlet clusters by comparing the topic vectors of their articles. We identified which topics are over-represented among articles from individual clusters by calculating the average topic vector of articles in the cluster and then comparing this vector to average topic vectors calculated after randomizing the political lean of each article's outlet. This null model, which we generate 1000 times, presents a random assignment of political label. For each topic, we calculated a *Z*-score comparing the prevalence of the topic in the observed articles with the average prevalence of the topic after randomization and scale by the standard deviation. We conducted this process separately for articles in each cluster.

LIWC features

Besides the significant differences in content between the derived clusters, we also examine important stylistic differences. This inquiry is driven by the observed effect of news story presentation on audience reception, especially along partisan lines. In particular, past work has demonstrated how stories that are subjective and convey emotion reinforce partisan isolation in news consumption and increase virality on social media (Berger and Milkman, 2012; Flaxman et al., 2016; Xu et al., 2020). Here, we investigate how these semantic aspects of news stories vary between our outlet clusters. We apply the Linguistic Inquiry and Word Count (LIWC) dictionary (Tausczik and Pennebaker, 2010), a widely used tool for investigating stylistic properties of text, to each article in our corpus. LIWC assigns text scores in various linguistic (e.g. the use of pronouns, prepositions, punctuation) and psychological (e.g. the use of words with significant positive or negative emotional valence) dimensions curated by human experts. LIWC features have been used to analyze the effectiveness of various kinds of persuasive writing, from crowdfunding pitches (Horvát et al., 2018) to public advocacy appeals (Bail et al., 2017). A study of democrats and republicans on Twitter using LIWC found significant differences between the characteristic linguistic style of the two groups (Sylwester and Purver, 2015), for instance, that republicans were more likely to use words expressing negative emotion.

Using the LIWC software, we scored each article in our corpus along 83 dimensions of linguistic features. We also considered two additional sentiment-based features from the VADER library (Hutto and Gilbert, 2014). We compared the distribution of each feature between the politically aligned clusters using a Mann–Whitney *U* test. We find that 50 out of 85 features have significant differences (Bonferroni corrected $p < .01$) in their distributions between the two groups of articles. The number of significant differences suggests substantial stylistic differences in the writing presented in the two clusters. The full table of features and differences are presented in Supplemental material Section 2.

Results

To address RQ1, we first investigated the patterns common across contributors' movements. Since movement occurs at the level of the news outlet, we analyzed what contributors' next steps tended to be from any given outlet (e.g. do contributors who write a story for the *New York Times* tend to also write for the *Washington Post*?). Figure 2(c) demonstrates these trends, by showing our outlet-to-outlet network with significant edges only. This network of statistically significant movements between news outlets reveals striking patterns in contributors' publishing histories. In particular, the following two clusters of news outlets emerge: One low-density cluster comprised nine outlets, and a dense cluster of four outlets. In the low-density cluster, only three outlets have connections with more than two outlets, creating a chain-like structure. From a contributor perspective, this indicates that jumps around the cluster are unlikely, but that travel to each outlet from every other is possible. In the high-density cluster, however, five of the six possible edges are present, indicating high mobility throughout the cluster. Between clusters, though, movement is highly unlikely. Outlets across clusters do not just lack significant positive connections; they have statistically significant negative edges. See the Supplemental material Table 1 for further details.

This clustering structure suggests that there are some constraints or barriers to the movement of contributors between outlets, in particular, as contributors seem to lack the flexibility to move between clusters. However, those constraints apply unevenly. Looking across contributor types, the majority of non-journalist authors, and academics cross between clusters at least once. However, less than a third of freelancers and other types of journalists move between clusters. For them, association with one cluster may represent a choice between two mutually-exclusive publishing paths, which have difficult-to-reverse consequences for future opportunities (Leung, 2014). In contrast, politicians make up 36% of all cross-cluster contributors, but only 5% of single-cluster contributors. Independent of the topics they focus on, then, politicians appear to move more freely between clusters.

These patterns are observable without taking into account any contributor or outlet attributes. Rather, they are structural characteristics that emerge from a combination of the editorial process and labor market. If the market structure of contributor-outlet relationships is built out of implicit patterns in activity, what heuristics—audience size or composition, for example—align with this structure? Are new classification mechanisms needed to understand the outlet landscape?

Structure of contributors' movements between news outlets

Traditionally, work that examines multiple outlets has divided them by their publishing medium—comparing print newspapers to news websites, for example, or radio to television (Maier, 2010). In this network, though, the outlet's traditional medium does not seem to align with the clusters we observe; magazine contributors also write pieces for radio and digital-native sites, for example, journalists shared by *The Atlantic, NPR*, and *Vox*. To validate this observation, we categorized each outlet using a variety of metadata, described in the "Data" section. Then, we measured how well each of these

classifications captures the clustering dynamic in our observed data using modularity (see details in "Methods" section). Our goal is to determine if traditionally accepted classification strategies can capture similar information about contributor movement patterns as our network structure, and if not, whether any other characteristics might provide a useful proxy.

In Figure 2(d), we show the modularity of each classification. Accordingly, outlet political leanings assigned by Bakshy et al. (2015) correspond best to the observed clustering. If we map the leanings of each outlet to our observed network (Figure 2(c)), we see a dense cluster of right-leaning outlets and a loose grouping of left- and center-leaning ones. Given this dichotomy, we subsequently refer to these clusters as left- or center-leaning and right-leaning. This finding suggests that the political leaning of a news outlet is a key structural factor in where contributors will publish, pointing, as in prior work, to a polarized online news ecosystem (Adamic and Glance, 2005; Bakshy et al., 2015; Benkler et al., 2018; Kaiser et al., 2019; Wihbey et al., 2019). While this structure impacts movement, though, does it affect the content of news articles across the network?

News coverage effects

The *Z*-scores derived from our topic model comparison across clusters (see "Methods" section) indicate the most relevant words of topics statistically over-represented among left- or center- and right-leaning outlets in the first columns of Table 2. A *Z*-score with absolute value greater than 1.96, corresponding to a *p*-value of .05, is considered a statistically significant deviation from the null hypothesis that cluster lean labels are unrelated to the distribution of topics. Articles appearing in center- or left-leaning outlets are significantly more likely to be about science and research (Topic 8), the media and press (Topic 3), and healthcare (Topic 16), among others. In right-leaning outlets, articles are more likely to be about conservatism and liberalism (Topic 10), prominent Democratic Party politicians (Topic 13), and Republican politics (Topic 4).

In total, 17 out of 20 topics are significantly over-represented among writings appearing in either the right- or left-leaning clusters of outlets. In relation to RQ2, this indicates that the partition of outlets according to transitions by contributors aligns with differences in content. Our findings prompt a question: Do those contributors that do transition between the left- and right-leaning clusters adapt their writing topics to the venue? Or do these transitioning contributors fill in particular niches within the partisan spheres they visit?

To address this question, consider three groups of contributors: those who write only for center- or left-leaning outlets, those who write only for right-leaning outlets, and those who write for both. Prior literature suggests that these groups will focus on different topics, for two reasons. First, if outlets gravitate toward certain political positions, they will often emphasize pieces of information relevant and favorable to that political group, and omit those that are not (Bernhardt et al., 2008). Therefore, if contributors write stories that are in line with the editorial approach of the outlet as a whole, we would also expect the information they produce to shift depending on their group of focus.

Second, it is challenging for either individual journalists or outlets as institutions to credibly maintain multiple political positions in their work (Anderson and McLaren,

Table 2. Topic keywords and their statistical over-representation (bold) and under-representation (italic) within news outlet clusters. The first two columns of Z-scores present statistical over or under-representation of topics within the center- or left- and right-leaning clusters. The last two columns record the statistical over or under-representations of topics in articles by contributors who move between clusters, compared to those by contributors who stay within the center- or left- and right-leaning clusters, respectively.

Topic: Keywords	CntrLeft Z	Right Z	Trans-CntrLeft Z	Trans-Right Z
1: attack, isi*, islam, muslim, war	0.42	0.8	−3.69	−5.29
2: vote, voter, percent, parti*, poll	−2.63	**4.07**	0.52	0.58
3: news, media, post, press, twitter	**9.99**	−9.97	−5.17	−1.16
4: senat*, cruz, candid*, gop, parti*	−5.62	**7.24**	**2.20**	**4.70**
5: realli*, talk, didn*, lot, someth*	**4.75**	−0.86	−5.55	0.38
6: game, play, team, season, serv*	**6.31**	−5.65	1.86	**6.85**
7: women, school, children, famili*, student	−1.04	**3.16**	−4.33	**5.62**
8: research, studi*, human, found, univers*	**13.07**	−12.37	−7.51	**2.18**
9: film, music, play, movi*, book	**5.23**	−3.47	−2.66	**3.37**
10: conserv*, power, left, liber*, fact	−23.59	**27.43**	**11.93**	−4.57
11: compani*, million, bank, busi*, billion	−6.64	−8.7	−3.24	**13.07**
12: court, law, immigr*, case, rule	−0.36	−0.76	**18.86**	−1.25
13: clinton, obama, hillari*, bill, email	−15.61	**14.45**	**2.3**	−6.63
14: white, black, christian, commun*, religi*	−3.3	**5.38**	−1.29	−2.86
15: tax, percent, job, econom*, rate	−3.92	**3.16**	−0.28	−0.91
16: health, care, plan, bill, insur*	**7.7**	−7.14	−3.64	−0.02
17: china, obama, unit, iran, foreign	**3.11**	−2.89	**6.92**	1.35
18: offici*, investig*, secur*, inform, depart	**6.2**	−6.21	−2.68	−5.58
19: citi*, build, water, climat*, area	**4.69**	−5.1	−1.54	−0.65
20: polic*, offic*, crime, gun, case	−2.06	**2.03**	−0.50	−4.51

2012). A contributor who writes a piece about the dangers of climate change, for example, cannot also credibly write a piece denying its existence. It is then in contributors' best interest to establish and develop a consistent perspective in their writing throughout their careers. The contributors who remain in only one cluster will generate that cluster's representative content (e.g. freelancers publishing in left-leaning outlets will focus on the media), while contributors who move between clusters will write about topics that are not particular to either. In other words, in response to RQ3, we expect that contributors moving between the left- and right-leaning outlets tend to write about politically more neutral topics.

We tested this idea by considering within-cluster differences. For both the left- and right-leaning cluster of outlets, we compared two kinds of contributors: those who stayed within the cluster, which we call purists, and those who transitioned. For example, among all articles in right-leaning outlets, we compare which topics tend to be covered by purists and those which tend to be covered by transitioning contributors. We repeated a similar experiment to the one used to determine topics over-represented among left- and right-leaning outlets. Within a cluster, we calculated the average topic vector of articles

Table 3. The top six LIWC features distinguishing texts from center- or left- (bold) and right-leaning (italic) outlets, by their area under the curve (AUC) scores.

Feature	M-W U	AUC	CntrLeft Avg.	Right Avg.	Diff.	Bonf. *P*
hear	3,144,691.0	0.65	1.08	0.71	0.37	$<10-89$
affect	3,242,798.0	0.64	4.33	5.10	−0.77	$<10-77$
percept	3,260,857.5	0.64	2.38	1.87	0.51	$<10-75$
negemo	3,438,952.5	0.62	1.88	2.36	−0.48	$<10-55$
certain	3,476,109.5	0.62	1.06	1.30	−0.24	$<10-51$
focuspast	3,616,885.0	0.60	4.09	3.45	0.64	$<10-38$

written by purists. We compared the average topic vector to those calculated from 1000 randomizations realized by shuffling the purist or transitioning contributor labels. This randomization allowed us to test the statistical over-representation of topics by purists or transitioning contributors within the partisan clusters.

The results of this investigation are shown in the last two columns of Table 2 for articles in left-leaning outlets and right-leaning outlets, respectively. Here, we report the relative prevalence of topics in articles written by transitioning authors within the two clusters, compared to their purist counterparts. In both clusters, there are many topics which are significantly over and underrepresented in the writings of transitioning authors. This indicates that switching contributors in both partisan clusters occupy specific niches with their writing.

Among the articles in right-leaning outlets, transitioning contributors are more likely to write about finance (Topic 11), sports (Topic 6), and families (Topic 7), among other topics. They are significantly less likely to write about Democratic Party politicians (Topic 13), investigations (Topic 18), and conflict in the Middle East (Topic 1). We note that of the six topics significantly underrepresented in the writings of transitioning contributors relative to purist, three topics are globally over-represented in right-leaning outlets (Topics 10, 13, 14). We interpret this as suggesting that transitioning contributors are less likely to write about extremely partisan topics.

The same analysis of left-leaning outlets suggests a similar pattern. Transitioning contributors writing for left-leaning outlets seem to avoid topics over-represented in those outlets on a global level (Topics 3, 8, 18). Two of the five topics in which switching contributors are over-represented are globally over-represented among right-leaning outlets. Future work should investigate whether this observed relationship is because transitioning contributors avoid partisan topics in order to remain widely employable, or perhaps because the kind of contributor who is able to switch between the partisan clusters tends to specialize in these topics.

Stylistic differences in language

Our LIWC and sentiment analysis also generate clear differences across outlet clusters. We report the top six most distinguishing features by effect size, measured by the area under the curve (AUC), in Table 3; the full table of features can be found in our

Supplemental material, Section 2. Three of these features are significantly over-represented in articles in center- and left-leaning outlets: *hear*, capturing the use of words describing the act of listening; *percept*, capturing the use of observational words relating to perception; and *focuspast*, capturing the use of past-tense verbs. Three are over-represented in right-leaning outlets: *affect*, counting the use of words with significant psychological content; *negemo*, counting the frequency of words with negative emotional connotation; and *certain*, counting the use absolute words such as "always" or "never."

At a high level, then, we observe widespread differences in the relative usage of certain linguistic and semantic markers across right- and center- or left-leaning outlets, giving us an affirmative answer to RQ2. Not only are audiences across these clusters exposed to different *content*; that content is also subject to differing *presentations*. In particular, we see that the usage of affect and emotion varies across these clusters, suggesting that the subjective frames with which writers present news stories depend on the preferences of the outlet and its audience.

At a high level, then, we observe widespread differences in the relative usage of certain linguistic and semantic markers across right- and center- or left-leaning outlets, giving us an affirmative answer to RQ2. Not only are audiences across these clusters exposed to different *content*; that content is also subject to differing *presentations*. In particular, we see that the usage of affect and emotion varies across these clusters, suggesting that the subjective frames with which writers present news stories depend on the preferences of the outlet and its audience.

Discussion

This work addresses an understudied area of news media polarization: structural production forces driving partisan leanings. By constructing a cross-outlet network purely based on contributor movement patterns, we show a clear partisan divide within the digital news ecosystem. That divide appears without any explicit consideration of audience behavior or preferences.

Our results demonstrate the ways in which structural factors can work against individual ones, such as a professional commitment to objectivity. Some contributors in our sample are not journalists, and some are explicitly political figures. However, those that do fall under the umbrella of professional journalism often stay within partisan bounds. Somewhere within the editorial process of pitching, selecting contributors, assigning stories, and producing news coverage, a dynamic arises that structurally prefers contributors whose publishing histories ideologically align with a publication's own. This may arise from institutional policies, from individual editors' preferences, or from the pitching process of individual contributors (Christin and Petre, 2020; Rosenkranz, 2018). In fact, the driving mechanisms may differ between outlets or individuals. These possible mechanisms warrant further study. Given the ways in which objectivity is so forcefully conveyed as a key standard at many outlets, it is also worth examining where outlet interactions with structural forces causes this norm to break down.

Polarized contributor isolation also shows up in the topic and style of news coverage they produce, wherein distinct areas of interest and modes of presentation arise for each outlet cluster. The polarization effect might originate with contributor movement, but it

also affects news coverage production at the story level. Prior work has demonstrated differences in news coverage approaches across the political spectrum (Schiffer, 2006; Xu et al., 2020). But the key here is that those differences do not exist in isolation, but are rather part of a larger media system that reinforces itself to drive polarization. Because of the integrated nature of polarization within this system, we cannot say what "causes" it. We can point to specific locations or mechanisms by which it appears, such as selective exposure or profit maximization at the outlet level (Anderson and McLaren, 2012; Stroud, 2010). However, saying any of those things *cause* polarization, or are even the main site of polarization, runs the risk of leading to pinpoint interventions that are broadly ineffectual. Audience-targeted interventions ignore producers' incentives to publish traffic-driving partisan coverage. Producer-targeted interventions ignore audiences' desires for similar perspectives to their own. Neither fully grapples with the feedback mechanism—running through the production, distribution, and consumption of digital news—tied to revenue incentives outlets must acknowledge to remain sustainable (Anderson and McLaren, 2012; McChesney, 2012). Future work should aim to work toward holistic interventions that address this integrated polarization.

We also find that topics less related to politics are more common for contributors who move between clusters. Relatively few contributors produce cross-cutting political news, reducing audiences' exposure to opposing viewpoints. Future work should examine whether more politically neutral topics are an effective starting point for cross-cutting exposure. In particular, it would be worthwhile to investigate whether the commonality of some topics also holds from an audience perspective—does sports coverage cut across political preferences, or does the partisan affiliation of a particular outlet affect all the coverage it produces? These results further highlight an interesting distinction between cross-cutting *topics* and *contributors*. Political topics are less common for cross-cluster contributors, but politicians make up a large share of those contributors. While politicians are still in the minority of all cross-cluster contributors (36%), their prominence raises questions about differing expectations and perceptions across types of contributors. These distinctions should be explored in future work.

Finally, in addition to addressing the broader area of news polarization research, this study also engages directly with structural production polarization work. Benkler et al. (2018) identify a very similar network structure in news media distribution to what we see here: a loosely connected group of left- or center-leaning outlets, a dense core of right-leaning outlets, and a very little activity between. Using a totally separate sample, and examining a different aspect of news production, we find an ecosystem with identical characteristics. Benkler et al. (2018) also demonstrate how the periphery of right-leaning outlets amplify conspiracy theories and misinformation into the core of mainstream media. In contrast, we see very little movement from this periphery outward, as contributors are significantly unlikely to move between clusters. This network structure is potentially good for stemming the flow of misinformation, but potentially bad for amplifying the content-based partisan leanings we see in both clusters. Guess and Coppock (2020) find that readers often update their views in response to information that contradicts their preconceptions. However, this updating process can only happen if readers are exposed to cross-cutting content in the first place, an unlikely prospect in a politically isolated media environment.

Limitations

This study has several important limitations. First, we cannot control for all the potential biases that may have arisen in data collection through scraping. We demonstrate the robustness of our results to differences in sample and significance level with additional checks. First, we check the robustness of the network to varying Z-score thresholds (Supplemental material, Section 1.1). We also examine the impact of removing individual outlets from the network on its structure (Supplemental material, Section 1.2). However, it is still important to acknowledge potential biases introduced by this data set.

Similarly, our sample represents a subset of all news outlets. While the outlets used here are prominent sources of digital media from across the political spectrum, they are only one part of the news ecosystem. On a larger scale, other factors may interact with polarization to provide a more complex picture of contributor movement. For example, prior work stresses the importance of considering the organizational context of news outlets when evaluating their coverage (Becker and Vlad, 2009). Factors such as the extent to which an outlet depends on freelancers, its reputation as a publishing venue, or its geographic focus may impact contributors' venue choices, and they may do so unevenly. This, in turn, might create an *unevenly* polarized network on a larger scale, affecting some outlets far more than others. To better understand how other factors might interact with, or even mitigate, the polarization we observe, future work should attempt to capture a broader subset of the news production ecosystem.

Finally, outlet political leaning is a difficult quality to measure. It originates from the perspective of the audience, rather than any objective measure of the news coverage an outlet produces. We use a rigorous, peer-reviewed method to ascertain our scores (Bakshy et al., 2015), but other approaches may categorize outlets differently.

Conclusion

This study demonstrates the polarized structure of digital news contributor movement. It ties that structure to differences in news coverage content and perspective across the political spectrum, and it highlights the topic differential across inter- and intra-cluster movement. Together, these results show the interconnected nature of polarization and its consequences across various aspects of news production. These findings fill an important gap in the literature on news media, highlighting structural factors on the production side which may contribute to polarization.

ORCID iD

Nick Hagar iD https://orcid.org/0000-0001-5110-3737

Supplemental material

Supplemental material for this article is available online.

Note

1. Because Reuters is a wire service from which stories often appear across many outlets, we could not distinguish true contributor movement from syndication. For this reason, we excluded Reuters from our sample.

References

Adamic L and Glance N (2005) The political blogosphere and the 2004 U.S. election: divided they blog. *LinkKDD '05*, pp. 36–43.

Anderson SP and McLaren J (2012) Media mergers and media bias with rational consumers. *Journal of the European Economic Association* 10(4): 831–859.

Bail CA, Argyle LP, Brown TW, et al. (2018) Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115(37): 9216–9221.

Bail CA, Brown TW and Mann M (2017) Channeling hearts and minds: advocacy organizations, cognitive-emotional currents, and public conversation. *American Sociological Review* 82(6): 1188–1213.

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.

Barnhurst KG and Mutz D (1997) American journalism and the decline in event-centered reporting. *Journal of Communication* 47(4): 27–53.

Becker LB and Vlad T (2009) News organizations and routines. In: Wahl-Jorgensen K and Hanitzsch T (eds) *The Handbook of Journalism Studies*. London: Taylor & Francis, pp. 59–72.

Benkler Y, Faris R and Roberts H (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Berger J and Milkman KL (2012) What makes online content viral? *Journal of Marketing Research* 49(2): 192–205.

Bernhardt D, Krasa S and Polborn M (2008) Political polarization and the electoral effects of media bias. *Journal of Public Economics* 92(5–6): 1092–1104.

Berton F, Devicienti F and Pacelli L (2011) Are temporary jobs a port of entry into permanent employment? Evidence from matched employer-employee. *International Journal of Manpower* 32(8): 879–899.

Blei DM, Ng AY and Jordan MI (2002) Latent Dirichlet allocation. In: *Advances in Neural Information Processing Systems*, pp. 601–608. Available at: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Blex C and Yasseri T (2020) Positive algorithmic bias cannot stop fragmentation in homophilic networks. *The Journal of Mathematical Sociology*. Epub ahead of print 28 June. DOI: 10.1080/0022250X.2020.1818078.

Boczkowski PJ and de Santos M (2007) When more media equals less news: patterns of content homogenization in Argentina's leading print and online newspapers. *Political Communication* 24(2): 167–180.

Bolsen T and Shapiro MA (2018) The US news media, polarization on climate change, and pathways to effective communication. *Environmental Communication* 12(2): 149–163.

Bruns A (2019) *Are Filter Bubbles Real?* Cambridge; Medford, MA: Polity Press.

Burt R (1992) *Structural Holes*. Cambridge, MA: Harvard University Press.

Chiricos T, Eschholz S and Gertz M (1997) Crime, news and fear of crime: toward an identification of audience effects. *Social Problems* 44(3): 342–357.

Christin A and Petre C (2020) Making peace with metrics: relational work in online news production. *Sociologica* 14: 133–156.

Cobb GW and Chen Y-P (2003) An application of Markov Chain Monte Carlo to community ecology. *The American Mathematical Monthly* 110(4): 265–288.

Cohen NS, Hunter A and O'Donnell P (2019) Bearing the burden of corporate restructuring: job loss and precarious employment in Canadian journalism. *Journalism Practice* 13(7): 817–833.

Coleman J (1988) Social capital in the creation of human capital. *American Journal of Sociology* 94: S95–S120.

Couldry N and Curran J (eds) (2003) *Contesting Media Power: Alternative Media in a Networked World* (Critical Media Studies). Lanham, MD: Rowman & Littlefield.

Dvir-Gvirsman S (2017) Media audience homophily: partisan websites, audience identity and polarization processes. *New Media & Society* 19(7): 1072–1091.

Flaxman S, Goel S and Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80(S1): 298–320.

Gionis A, Mannila H, Mielikäinen T, et al. (2007) Assessing data mining results via swap randomization. *ACM Transactions on Knowledge Discovery from Data* 1(3): 14.

Gollmitzer M (2014) Precariously employed watchdogs? Perceptions of working conditions among freelancers and interns. *Journalism Practice* 8(6): 826–841.

Granovetter MS (1973) The strength of weak ties. *American Journal of Sociology* 78(6): 1360–1380.

Guess A and Coppock A (2020) Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science* 50(4): 1497–1515.

Harder RA, Sevenans J and Van Aelst P (2017) Intermedia agenda setting in the social media age: how traditional players dominate the news agenda in election times. *The International Journal of Press/Politics* 22(3): 275–293.

Hayes K and Silke H (2018) The networked freelancer? Digital labour and freelance journalism in the age of social media. *Digital Journalism* 6(8): 1018–1028.

Heatherly KA, Lu Y and Lee JK (2017) Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media & Society* 19(8): 1271–1289.

Helberger N (2019) On the democratic role of news recommenders. *Digital Journalism* 7(8): 993–1012.

Hindman M (2018) *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. Princeton, NJ: Princeton University Press.

Hopkins DJ and King G (2010) A method of automated nonparametric content analysis for social science. *American Journal of Political Science* 54(1): 229–247.

Horvát E-A, Wachs J, Wang R, et al. (2018) The role of novelty in securing investors' for equity crowdfunding campaigns. In: *Sixth AAAI Conference on Human Computation and Crowdsourcing*. Available at: https://www.aaai.org/ocs/index.php/HCOMP/HCOMP18/paper/viewFile/17929/16909

Hutto CJ and Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth International AAAI Conference on Weblogs and Social Media*. Available at: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Kaiser J, Rauchfleisch A and Bourassa N (2019) Connecting the (far-)right dots: a topic modeling and hyperlink analysis of (far-)right media coverage during the US elections 2016. *Digital Journalism* 8(3): 422–441.

Knobloch-Westerwick S and Meng J (2009) Looking the other way: selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research* 36(3): 426–448.

Leung MD (2014) Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *American Sociological Review* 79(1): 136–158.

Lowrey W (2006) Mapping the journalism–blogging relationship. *Journalism: Theory, Practice & Criticism* 7(4): 477–500.

McCallum AK (2002) Mallet: a machine learning for language toolkit. *Technical report, UMass*. Available at: http://mallet.cs.umass.edu

McChesney RW (2003) The Problem of Journalism: a political economic contribution to an explanation of the crisis in contemporary US journalism. *Journalism Studies* 4(3): 299–329.

McChesney RW (2012) FAREWELL TO JOURNALISM? Time for a rethinking. *Journalism Studies* 13(5–6): 682–694.

McCombs ME and Winter JP (1981) Defining local news. *Newspaper Research Journal* 3(1): 16–21.

Macy M, Deri S, Ruch A, et al. (2019) Opinion cascades and the unpredictability of partisan polarization. *Science Advances* 5(8): eaax0754.

Maier D, Waldherr A, Miltner P, et al. (2018) Applying lda topic modeling in communication research: toward a valid and reliable methodology. *Communication Methods and Measures* 12(2–3): 93–118.

Maier S (2010) All the news fit to post? Comparing news content on the web to newspapers, television, and radio. *Journalism & Mass Communication Quarterly* 87(3–4): 548–562.

Mitchelstein E and Boczkowski PJ (2009) Between tradition and change: a review of recent research on online news production. *Journalism: Theory, Practice & Criticism* 10(5): 562–586.

Mohr JW and Bogdanov P (2013) Introduction—topic models: what they are and why they matter. *Poetics* 41(6): 545–569.

Montgomery JM and Nyhan B (2017) The effects of congressional staff networks in the US house of representatives. *The Journal of Politics* 79(3): 745–761.

Mullainathan S and Shleifer A (2005) The market for news. *American Economic Review* 95(4): 1031–1053.

Munger K (2020) All the news that's fit to click: the economics of clickbait media. *Political Communication* 37(3): 376–397.

Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E* 64(1): 016131.

Newman MEJ (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23): 8577–8582.

Newman MEJ (2010) *Networks: an Introduction*. Oxford: Oxford University Press.

Rao AR, Jana R and Bandyopadhyay S (1996) A Markov Chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā: The Indian Journal of Statistics, Series A* 58(2): 225–242.

Řehůřek R and Sojka P (2010) Software framework for topic modelling with large corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta: ELRA, pp. 45–50. Available at: http://is.muni.cz/publication/884893/en

Resnick P, Garrett RK, Kriplean T, et al. (2013) Bursting your (filter) bubble: strategies for promoting diverse exposure. In: *Proceedings of the 2013 Conference on Computer Supported*

Cooperative Work Companion—CSCW '13. ACM Press, p. 95. Available at: http://dl.acm.org/citation.cfm?doid=2441955.2441981

Ribeiro MH, Ottoni R, West R, et al. (2020) Auditing radicalization pathways on YouTube. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Barcelona: ACM Press, pp. 131–141. Available at: https://arxiv.org/abs/1908.08313

Rosenkranz T (2018) From contract to speculation: new relations of work and production in freelance travel journalism. *Work, Employment and Society* 33(4): 613–630.

Ryan M (2001) Journalistic ethics, objectivity, existential journalism, standpoint epistemology, and public journalism. *Journal of Mass Media Ethics* 16(1): 3–22.

Salamon E (2020) Digitizing freelance media labor: a class of workers negotiates entrepreneurialism and activism. *New Media & Society* 22(1): 105–122.

Schiffer AJ (2006) Assessing Partisan bias in political news: the case(s) of local senate election coverage. *Political Communication* 23(1): 23–39.

Schlauch WE, Horvát EA and Zweig KA (2015) Different flavors of randomness: comparing random graph models with fixed degree sequences. *Social Network Analysis and Mining* 5: 36.

Schudson M (2001) The objectivity norm in American journalism. *Journalism: Theory, Practice & Criticism* 2(2): 149–170.

Schudson M and Anderson C (2009) Objectivity, professionalism, and truth seeking in journalism. In: Wahl-Jorgensen K and Hanzitsch T (eds) *The Handbook of Journalism Studies*. New York: Routledge, pp. 88–101.

Shi F, Shi Y, Dokshin FA, et al. (2017) Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour* 1(4): 0079.

Singer JB (1997) Still guarding the gate? The newspaper journalist's role in an online world. *Convergence: The International Journal of Research into New Media Technologies* 3(1): 72–89.

Singer JB (2005) The political j-blogger: "Normalizing" a new media form to fit old norms and practices. *Journalism: Theory, Practice & Criticism* 6(2): 173–198.

Stroud NJ (2010) Polarization and partisan selective exposure. *Journal of Communication* 60(3): 556–576.

Sylwester K and Purver M (2015) Twitter language use reflects psychological differences between democrats and republicans. *PLoS ONE* 10(9): e0137422.

Taneja H and Wu AX (2018) Pathways to fragmentation: user flows and web distribution infrastructures. In: *Proceedings of the 10th ACM Conference on Web Science (WebSci'18)*. Amsterdam: ACM Press, pp. 255–259. Available at: https://arxiv.org/abs/1803.02331

Tausczik YR and Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1): 24–54.

Tewksbury D and Riles JM (2015) Polarization as a function of citizen predispositions and exposure to news on the Internet. *Journal of Broadcasting & Electronic Media* 59(3): 381–398.

Thompson A (2017) All the news:143,000 articles from 15 American publications. Available at: https://www.kaggle.com/snapcrack/all-the-news

Tsfati Y, Stroud NJ and Chotiner A (2014) Exposure to ideological news and perceived opinion climate: testing the media effects component of spiral-of-silence in a fragmented media landscape. *The International Journal of Press/politics* 19(1): 3–23.

Tumminello M, Miccichè S, Lillo F, et al. (2011) Statistically validated networks in bipartite complex systems. *PLoS ONE* 6(3): e17994.

Vedres B and Stark D (2010) Structural folds: generative disruption in overlapping groups. *American Journal of Sociology* 115(4): 1150–1190.

Webster JG (2016) The marketplace of attention. In: Webster JG (ed.) The Marketplace of Attention: How Audiences Take Shape in a Digital Age. Cambridge, MA; London: The MIT Press, pp. 1–22.

Wihbey J, Joseph K and Lazer D (2019) The social silos of journalism? Twitter, news media and partisan segregation. *New Media & Society* 21(4): 815–835.

Xu WW, Sang Y and Kim C (2020) What drives hyper-partisan news sharing: exploring the role of source, style, and content. *Digital Journalism* 8(4): 486–505.

Zweig KA and Kaufmann M (2011) A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining* 1(3): 187–218.

Author biographies

Nick Hagar is a PhD candidate in the Computational Journalism Lab at Northwestern University. His research focuses on collective attention problems in digital media.

Johannes Wachs is an assistant professor at the Institute for Data, Process and Knowledge Management of the Vienna University of Economics and Business (WU Wien). He is also affiliated with the Complexity Science Hub. He studies social, technical, and economic networks to understand their effects on people and society.

Emőke-Ágnes Horvát is an Assistant Professor in the Department of Communication Studies at Northwestern University. Her research at the intersection of computational social science and social computing develops network and big data methods to understand and support collective intelligence in Web-based systems.