

# Editorial Versus Audience Gatekeeping: Analyzing News Selection and Consumption Dynamics in Online News Media

Abhijnan Chakraborty<sup>1</sup>, Saptarshi Ghosh<sup>2</sup>, Niloy Ganguly, *Senior Member, IEEE*, and Krishna P. Gummadi

**Abstract**—In recent years, we have witnessed a paradigm shift in news consumption. In traditional news media organizations, a small number of expert editors are responsible for selecting news stories that are consumed by all news readers (the audience). However, with the growing popularity of social media as a news consumption medium, a part of the editorial power of selecting news stories has shifted to the audience who select and share the stories that can reach a large number of consumers. In this paper, we analyze data from two popular news media sites—The New York Times and The Guardian, and characterize the considerable differences in the types of stories selected by the audience and expert news editors. We also find that story selections by audience vary significantly across different social media channels such as Twitter, Facebook, and email. We contextualize the differences utilizing media and communication theory and discuss their implications for news readers and media organizations.

**Index Terms**—Editorial selection, journalistic gatekeeping, online news consumption, social news sharing.

## I. INTRODUCTION

JOURNALISTS in different news organizations have long seen their occupational roles as the *gatekeepers of information* determining what news should be made available to the public [1]. Such a gatekeeping role also included highlighting certain news stories more prominently than others, e.g., placing some of the stories on the front page of the printed newspaper. Such a differential treatment of certain stories provides cues to the readers about their importance and allows the news organizations to set the *news agenda* for the broader society [2]. Of course, a news reader may not allocate one’s attention equally to all news stories made available by the gatekeepers. They may show preference toward some stories more than others

and abandon some of the stories altogether—a phenomenon termed as *selective exposure* [3], whereby readers choose to expose themselves more to news that is consistent with their standpoints [4].

In recent times, such a mismatch between the readers’ and the editors’ choice is getting magnified with the widespread adoption of social media, where a large and growing fraction of news readers are finding (or being led to) news stories over platforms such as Facebook and Twitter [5]. Such a shift has prompted news websites to add one-click sharing buttons to the news stories, using which a member of the audience can easily share a news story over social media like Twitter, Facebook, or email. Such social feedback enables different stories to reach a potentially large number of people. Thus, today, the news media audience also act as *network gatekeepers* [6], who selectively decide which stories should reach their peers.

Prior works in media studies have looked into the coverage of news across such different selections [2], [7], [8]. However, these works have focused either on the editors’ selection versus selective exposure by the audience or the network gatekeeping across social networks. To the best of our knowledge, there has been no work that considered all *three different dynamics* (editors’ selection, readers’ consumption, and readers’ sharing pattern) together undergoing in the current media landscape. In this paper, we attempt to systematically understand the differences between the selections performed by expert editors and the audience, by comparing news stories they select (either implicitly for viewing or explicitly for sharing) from the same candidate set of stories.

To this end, we collected all news stories published on the popular news media site `nytimes.com` (NYTimes) over a period of 8 months. During this period, we gathered stories recommended by NYTimes expert editors, as well as the stories most popularly viewed and shared by the audience over Facebook, Twitter, and email. We then introduce a novel framework to compare the news coverage of different news selections, and using this framework, we compare expert versus audience selection as well as audience selections for sharing across different social media.

We find that there are significant differences in the coverage of news stories selected by editors versus the audience. For example, “World,” “Sports,” and “Business” stories are more selected by the experts than the audience. On the other hand, “Opinion” pieces as well as stories on “Science,” “Fashion,” and “Health” are selected by the audience much more than by the experts. We also find significant differences in news

Manuscript received April 15, 2018; revised September 7, 2018, February 18, 2019, and April 11, 2019; accepted May 23, 2019. Date of publication June 25, 2019; date of current version August 8, 2019. This work was supported in part by the European Research Council (ERC) Advanced Grant for the project “Foundations for Fair Social Computing,” through the European Union’s Horizon 2020 Framework Program under Grant 789373. The work of A. Chakraborty was supported by the Google India Ph.D. Fellowship and the Prime Minister’s Fellowship for Doctoral Research, a public-private partnership between Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, and Confederation of Indian Industry (CII). (*Corresponding author: Abhijnan Chakraborty.*)

A. Chakraborty was with IIT Kharagpur, Kharagpur 721302, India. He is now with the Max Planck Institute for Software Systems, 66123 Saarbrücken, Germany (e-mail: achakrab@mpi-sws.org).

S. Ghosh and N. Ganguly are with IIT Kharagpur, Kharagpur 721302, India. K. P. Gummadi is with the Max Planck Institute for Software Systems, 66123 Saarbrücken, Germany.

Digital Object Identifier 10.1109/TCSS.2019.2920000

consumption and sharing by the audience. Political stories and stories related to law and crime are consumed more but shared less. On the other hand, “Health,” “Travel” stories, and “Opinion” pieces are shared more. Our analysis further reveals significant differences in the type of news shared by the audience over different social media. For example, “Opinion” pieces and local stories (related to the sections “New York” and “U.S.”) tend to be more popularly shared on Facebook, while news stories from “Business” and “World” sections are more shared on Twitter.

To validate whether the above-mentioned observations generalize beyond NYTimes stories, we analyzed news stories from another popular media site—The Guardian. We found that our observations are qualitatively very similar for both news media sites. Thus, the insights obtained from this paper are generalizable across different media sites (potentially having different editorial strategies). We further contextualize the differences utilizing prior works in media and communication studies and discuss the implications of our findings for the news readers and the media organizations.

## II. DATA SET AND METHODOLOGY

As mentioned earlier, in this paper, our objective is to analyze the coverage of news stories selected by the expert editors and the audience for sharing in different media. We attempt to fulfill this objective in the context of a popular news media site—`nytimes.com` (NYTimes).

### A. Data Set Gathered

We undertook a continuous data collection drive over a period of 8 months: July 2015–February 2016. We collected 191 807 distinct news stories appearing on NYTimes during this period, using the Article Search API [9]. At any point, NYTimes editors recommend around 20 stories at the top of the homepage [10] (thus termed as “Top Stories”). We collected these top stories by querying the Top Stories API [11] at every 5-min intervals throughout the 8-month period.

To facilitate social sharing of news stories, NYTimes provides *one-click sharing buttons* just under the headline of every story, and tracks the news stories getting most shares in different media. Via the Most Popular API [12], it allows programmatic access to 20 stories most viewed, most emailed, most tweeted, and most shared on Facebook—all of which are computed over the last 24 hours from the time of access.<sup>1</sup> We collected all these stories by querying the API at every 5-min intervals throughout July 2015–February 2016. Table I(a) gives the number of distinct stories that appeared in different sets during this period. NYTimes Article Search API also provides detailed metadata regarding every news story (e.g., its author, headline, summary, section, and the topics assigned to it by NYTimes). We gathered these metadata for all stories, which we use extensively in our analysis.

Effectively, top stories represent the *editorial gatekeeping* in the selection of news stories, and most viewed stories

<sup>1</sup>Even if a particular reader views or shares the same news story multiple times, the API considers it once. Also, accidental clicks (i.e., very short duration stays) do not count while computing the list of most viewed stories.

TABLE I

(A) NUMBER OF DISTINCT NYTIMES NEWS STORIES THAT GOT SELECTED EITHER BY EXPERT EDITORS OR BY THE AUDIENCE FOR VIEWING AS WELL AS SHARING VIA DIFFERENT NEWS DISSEMINATION MEDIA. (B) NUMBER OF DISTINCT TOPICS COVERED BY SUCH STORIES DURING JULY 2015–FEBRUARY 2016

Type	(a) Distinct Stories	(b) Topics Covered
All Stories Published	191,807	41,329
Top Stories	10,348	10,213
Most Viewed Stories	3,970	5,239
Most Shared Stories Across All Channels	3,537	5,913
Most Emailed Stories	3,114	5,661
Most Tweeted Stories	2,799	4,500
Most Shared Stories on Facebook	2,771	4,157

represent the *selective exposure* of news audience for consumption, while most shared stories across different media (e.g., email, Twitter, and Facebook) represent the *networked gatekeeping* performed by the crowd in selecting news stories for sharing. Hence, choosing the data available from NYTimes enables us to gather reliable, large-scale data on selections of news stories by the expert editors and audience, on the *same set of news stories*.

### B. Comparing News Coverage by Different Selections

We compare the news coverage by different selections of stories at three levels: 1) coverage of individual news stories; 2) sectional coverage; and 3) topical coverage.

1) *Comparing Individual Stories Covered*: While comparing the coverage at the level of individual news stories, we compute the overlap (and differences) between the *set* of all stories selected during the entire measurement period. For a visual representation of the result, we use *Venn diagrams* that pictorially demonstrate the comparisons in terms of the absolute number of news stories. One important thing to note that in this analysis, we only consider *unique* stories, i.e., even if a story is selected (e.g., by editors) multiple times during our measurement period, we consider it once.

2) *Comparing Sectional Coverage*: To help the readers easily navigate through the vast collection of news stories published, news organizations (e.g., NYTimes) assign a story to a particular *news section* such as “Business” or “Sports.” We measure the sectional coverage of a collection of stories as the *distribution of stories over these news sections*. Specifically, if  $x_i$  fraction of stories are assigned to section  $i$ , and there are  $m$  news sections in a media website, then the sectional coverage is computed as the distribution of these fraction values  $\{x_1, x_2, \dots, x_m\}$  over all  $m$  news sections. Table II lists all news sections at NYTimes, and hence, the sectional coverage is computed over them.

To compare the sectional coverage of two different selections, we first compute the sectional coverage of individual lists of 20 selected stories (i.e., the list of 20 stories returned by the API at any particular time instant) and repeat this procedure over all such lists gathered during the data collection period. To compare two selection strategies, we consider each

TABLE II  
SECTIONS ASSIGNED TO NYTIMES STORIES

News Sections
Automobiles, Arts, Books, Business Day, Education, Fashion & Style, Food, Health, Magazine, Movies, New York, Obituaries, Opinion, Politics, Real Estate, Science, Sports, Sunday Review, Technology, T Magazine, Television, The Upshot, Theater, Times Insider, Travel, U.S., World

section  $i$ , and for both strategies, compute the *mean* value  $\hat{x}_i$  from the  $x_i$ s calculated for individual lists. Then, we run *Welch's T-Test*<sup>2</sup> [14] to check whether the differences between the two  $\hat{x}_i$ s are *statistically significant*. If for section  $i$ , the  $p$ -value observed in the t-test is less than the *significance level*  $\alpha$  (we set  $\alpha = 0.05$ ), we conclude that the difference in coverage of the section  $i$  is statistically significant. We repeat this process to find all sections with significant differences in sectional coverage.

To compare more than two (say  $n$ ) selections, we perform  $c = \binom{n}{2}$  pairwise t-tests, after applying *Bonferroni Correction* [15] to lower significance level to  $(\alpha)/(c - 1)$ . We conclude the presence of statistically significant difference in the coverage of section  $i$  only when the  $p$ -value is below  $(\alpha)/(c - 1)$  in such comparisons.

3) *Comparing Topical Coverage*: To better organize different news stories around specific themes, NYTimes assigns around five topics (or keywords) to every news story. Such topics help in the organization as well as in search and retrieval of news stories. For example, the NYTimes page on the topic “Guns and Gun Control”<sup>3</sup> shows the collection of all news stories related to that topic. NYTimes deploys a combination of manual and algorithmic techniques to assign these topics [16]. Table I(b) gives the number of distinct topics assigned to the stories that got selected by the editors and the audience.

For comparing the topical coverage of two selections, we analyze the topics that are covered more *prominently* in one and not in the other. To find such topics, we first compute how frequently different topics are covered by a particular selection strategy, and then for each strategy, we sort the topics based on their frequencies, where we measure the *frequency*  $y_j$  as the number of news stories covering topic  $j$ . From the sorted list of topics, we can find the rank of any topic. For example, a particular topic  $t$  may have rank  $r_1$  for strategy  $S_1$  and rank  $r_2$  for strategy  $S_2$ . We conclude that the topic is more prominently covered by  $S_1$ , if the difference between  $r_1$  and  $r_2$  is *statistically significant* according to *Fisher's exact test* [17], and  $r_1 < r_2$ .

Specifically, if the number of distinct topics covered by  $S_1$  and  $S_2$  is  $N_1$  and  $N_2$ , respectively, then to evaluate whether

the difference in ranks  $r_1$  and  $r_2$  for the topic  $t$  is significant, we build the following *Contingency Table* [17]

$$\begin{vmatrix} r_1 - 1 & N_1 - r_1 \\ r_2 - 1 & N_2 - r_2 \end{vmatrix}.$$

To illustrate the construction of the contingency table, assume that a particular topic  $t$  is ranked 25th out of 200 topics covered by  $S_1$ , and ranked 170th out of 250 topics covered by  $S_2$ . Then, the contingency table will become

$$\begin{vmatrix} 24 & 175 \\ 169 & 80 \end{vmatrix}.$$

We compute the  $p$ -value using this contingency table.<sup>4</sup> If the  $p$ -value is less than 0.05, we conclude that the difference between  $r_1$  and  $r_2$  is statistically significant. We repeat this procedure for all topics covered by the strategies  $S_1$  and  $S_2$ . To account for topics covered by only one strategy (say  $S_1$ ), we assume that they are ranked last in the other strategy (i.e.,  $r_2 = N_2$ ). This way, Fisher's exact test can be carried out uniformly for all topics.

*Coverage of Political Topics*: NYTimes categorizes the topics into four types: *Subject* (e.g., “Presidential Election of 2016” and “Books and Literature”), *Location* (e.g., “New York City” and “Syria”), *Person* (e.g., “Donald J Trump,” and “Barack Obama”), and *Organization* (“House of Representatives” and “New York Mets”). To further characterize the persons covered by the NYTimes stories, we utilized Wikipedia to identify their political affiliations. For most of the people, we could find their corresponding Wikipedia pages. We collected information about their political party if it was part of their personal details in Wikipedia. As NYTimes stories cover both national and international political persons, the political parties covered also include parties from other countries (e.g., Indian National Congress and Communist Party of China). In this paper, we restrict ourselves to two major political parties in the U.S.—Republican Party and Democratic Party and only consider the coverage of people affiliated to these two parties.

### III. ANALYSIS

Using the measures described in Section II, we first compare the coverage of news stories selected by the expert editors and the audience and then stories selected by the audience for sharing across different media platforms (e.g., email, Facebook, or Twitter).

#### A. Comparing News Selection by the Editors Versus the Audience

As editor-selected stories, we consider the top stories on NYTimes, while as audience-selected stories, we consider

<sup>2</sup>Welch's T-Test is preferred over more popular Student's T-Test when the two samples being compared have unequal variances and unequal sample sizes [13]. In our context, different selections cover a particular section  $i$  in different manners at different times, resulting in unequal variances in the coverage of  $i$ . Moreover, there are times when only one of the selections cover  $i$  and the other one may not cover  $i$  at all, resulting in unequal sample sizes. To account for these two factors, we used Welch's T-Test in our experiments.

<sup>3</sup>[nytimes.com/topic/subject/guns-and-gun-control](http://nytimes.com/topic/subject/guns-and-gun-control)

<sup>4</sup>Fisher's exact test is one of the most popular statistical tests on  $2 \times 2$  contingency tables [18]. The most common use case for this test is categorical data resulting from the classification of the same object in two different manners. The test directly outputs the  $p$ -value denoting the significance of the association between two types of classification. In our context, two selections are covering the same topic in a different frequency, resulting in their different categorical ranks. The contingency table shows the association between these two types of rankings. We apply Fisher's exact test to find the topics where the rankings are significantly different.



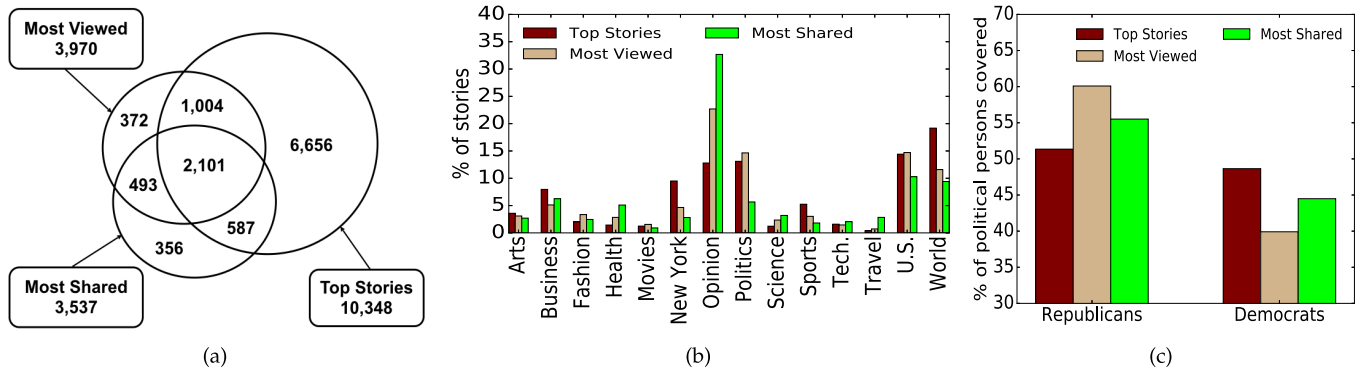


Fig. 1. Comparing the editor-selected top stories with the audience-selected most viewed and most shared stories, according to (a) overlap of individual stories, (b) their sectional coverage, and (c) coverage of political persons.

most viewed stories, representing the selective exposure of the audience, and most shared stories, representing their network gatekeeping role. Note that sharing is more involved way of audience selection because it involves two steps: first, they need to select the news stories for consumption on the news platform and then select a subset of these stories for sharing.

1) *Difference in Individual News Stories:* We compute the overlap and difference in the set of top stories, most viewed, and most shared stories and present the statistics in Fig. 1(a). We observe that there are significant fractions from each set of stories, that are *not* covered by the other set. For example, between top stories and most viewed stories, being a smaller set, it is expected that most viewed stories cannot fully cover the larger set of top stories. However, the interesting finding from Fig. 1(a) is that *despite being three times larger than most viewed stories, top stories do not cover 22% (372 + 493 = 865 out of 3970) of the most viewed stories.* Similarly, 24% (356 + 493 = 849 out of 3537) most shared stories are not part of the top stories selected by expert editors. Surprisingly, we find that audience selection also differs significantly between consumption and sharing, where 35% (372 + 1004 = 1376 out of 3970) of most viewed stories are not most shared. Similarly, 27% (356 + 587 = 943 out of 3537) of the most shared stories are not most viewed.

Differences between the editor and audience selections need to be contextualized with respect to the prior works on the importance of news story placement. Past research works have shown that the placement of news stories in highly visible places offer *story importance cues* to the readers to influence them to interpret the stories as being important [2]. Editor-selected top stories are promoted actively on NYTimes homepage offering enough importance cues to the readers. Yet, 22% of the most viewed stories (similarly 24% of most shared stories), which had no associated importance cues, got exclusively picked by the audience. This clearly indicates that the news audience carve their own paths in choosing which stories to select, thereby abandoning the sole reliance on expert editorial control.

We attempt to understand the difference between consumption and sharing behavior of the audience using prior works on the communication theory. Researchers found that one of the main motivations for sharing information is to get social

recognition [19]. The person sharing news can appear well informed, and the person receiving it may recognize the same. However, the content of the news determines whether a person would be willing to be recognized in certain ways. Harber and Cohen [20] found that people tend to share *bad news* (e.g., stories on disasters, accidents, and death of celebrities) that they feel emotionally connected to. On the other hand, people tend to avoid sharing controversial stories, particularly when the intended audience is known to hold contrary opinion [21]. Similarly, people do not share certain kinds of stories they want to consume in private (i.e., “guilty pleasure” [22]) but do not want to get associated with. For these reasons, we observe that many stories, which are selected by the audience for consumption, do not get shared often; whereas, some stories that are not much consumed, get shared by the audience to maintain their well-informed persona.

2) *Difference in Sectional Coverage:* To further characterize the differences in choices of the editors and the news audience, we compare the sectional coverage of top, most viewed, and most shared stories, and Fig. 1(b) shows the result. We run 3 two-sample T-tests for every section and found statistically significant differences in coverage of sections in Fig. 1(b) (at significance level 0.025, after applying the Bonferroni Correction [15]). We observe the following trends in Fig. 1(b).

- 1) NYTimes stories on certain sections of broad interest like Politics and U.S. have similar coverage in the top and most viewed stories, while those on other sections of broad interest like World, Sports, and Business are much less consumed by the audience (compared to their coverage in editor selections).
- 2) For stories across all sections of broad interest, audience shares them much less compared to what editor selects for top stories and consumed by the audience.
- 3) Opinion pieces and niche interest stories on Science, Travel, and Health are shared by the audience more than their fraction in editor-selected stories.

The differences between the top and most viewed stories are possibly due to the audience associating NYTimes more with political and national news and news on several niche soft categories, but less with international news or news related to sports or business. However, with respect to sharing, the audience tend to share only news (for example, opinion

TABLE III  
TOPICS PROMINENTLY COVERED BY EITHER THE EDITOR SELECTED TOP STORIES OR AUDIENCE SELECTED MOST VIEWED STORIES

Topic Type	Example topics prominently covered by the editors, but not by the audience
Subject	Law and Legislation, Suits and Litigation (Civil), Corruption (Institutional), Embargoes and Sanctions, Human Rights & Human Rights Violations, Football, Baseball, Automobile Safety Features & Defects
Location	Los Angeles (Calif), Baltimore (MD), Cuba, Pakistan, West Bank, Taiwan, Ferguson (MO), Caribbean
Person	Andrew M Cuomo, John Kerry, Ashton B Carter, Narendra Modi, Bill Cosby, Kim Jong-un
Organization	Police Department (NYC), Communist Party of China, Central Intelligence Agency, Afghan National Security Forces, Kurdistan Workers' Party, New York Giants, Carolina Panthers, New York Jets
Topic type	Example topics prominently covered by the audience, but not by the editors
Subject	Constitution (US), Social Security (US), Space and Astronomy, Diet and Nutrition, Dating and Relationships, Exercise, Religion and Belief, Ethics (Personal), Weddings and Engagements
Location	San Bernardino (Calif), Hungary, Brussels, Indiana, East Village (NY), Midtown Area (NY), Reunion Island, Jersey City
Person	Justin Trudeau, George Bush, Michelle Obama, Adele, Roger Federer, Beyonce, Kate McKinnon
Organization	National Broadcasting Co, CNBC, National Oceanic and Atmospheric Administration, Kennedy International Airport (Queens, NY), New York Public Library, Playboy Magazine, Starbucks

TABLE IV  
TOPICS PROMINENTLY COVERED BY EITHER MOST VIEWED STORIES OR MOST SHARED STORIES

Topic type	Example topics prominently covered for consumption, but not for sharing
Subject	Debates (Political), Primaries and Caucuses, Murders, Attempted Murders and Homicides, Polls and Public Opinion, Sex Crimes, School Shootings and Armed Attacks, Muslim Americans
Location	New Hampshire, Nevada, New Jersey, Colorado Springs, Virginia, North Korea
Person	Bill de Blasio, Bobby Jindal, Stephen Colbert, Huma Abedin, John McCain, Michael R Bloomberg
Organization	Fox News, CNN, Taliban, IMF, Federal Election Commission, Republican National Committee
Topic type	Example topics prominently covered for sharing, but not for consumption
Subject	Travel and Vacations, Books and Literature, Global Warming, Social Conditions and Trends, Education (K-12), Income Inequality, Medicine and Health, Agriculture and Farming
Location	Cuba, Africa, Italy, Latin America, Arctic Regions, Caribbean Area, Seattle (Wash), Atlanta (Ga)
Person	Charles Koch, Harper Lee, Martin Winterkorn, Timothy D Cook, Cory Booker, Bill O'Reilly
Organization	Apple Inc, Food and Drug Administration, Homeland Security Department, Journal of American Medical Association, Tesla Motors Inc

pieces) that are exclusively available at NYTimes, sharing which such readers can play the roles of *opinion leaders* [23], and initiate discussions in their communities around the news stories. Similarly, the audience feel emotionally connected to stories on Health or Travel and wants to share that emotion among their peers [20]. On the other hand, they tend to avoid sharing political and international news.

3) *Difference in Topical Coverage*: To compare the topical coverage of the editor and audience selections, we find the topics covered more prominently in one selection compared to the other. Table III gives the example topics that are prominently selected either by the editors or the audience. We see in Table III that editors play a larger role of *journalistic gatekeeping* [1] by promoting more *hard* stories related to law and crime, international events, or societal impacts of government policies, what they think are important for the readers to know. The audience, on the other hand, along with topics related to public affairs, also pays attention to *soft* human interest stories (such as entertainment or leisure) more than what the editors select.

Table IV presents the topics selected by audience either for consumption or for sharing. Among the topics selected for consumption, we find topics related to election debates, law, and order situation or the controversies around the role of different media houses (e.g., CNN and Fox News) during the

election period. On the other hand, topics prominently selected for sharing include topics related to business, arts, literature, health, or environmental issues.

Finally, to investigate how people from different political parties are covered, we plot in Fig. 1(c), the fraction of Republican and Democrat politicians covered by different selections (as mentioned earlier, we do not consider other political parties and normalize the fractions accordingly). Interestingly, we observe that although NYTimes is perceived to have a *liberal bias* [24], the editor selected top stories covered both republican and democrat politicians almost equally. However, the audience consumed more stories covering republicans than stories covering democrats. While sharing, the difference between fractions reduced, but still the coverage of republicans, was more than democrats.

4) *Summary*: We observed that there exists a gap between the newsroom decisions and their audience, where the audience selectively consumes stories that differ significantly from the stories promoted by the news organizations, and thereby reducing the power of the media organizations to set the news agenda. One could debate whether such a situation is good or bad. On the one hand, news selection by audience offer the democratization of news curation, breaking the agenda-setting power of a coterie of news editors. However, the recent debate on the impact of social media in the outcome of elections

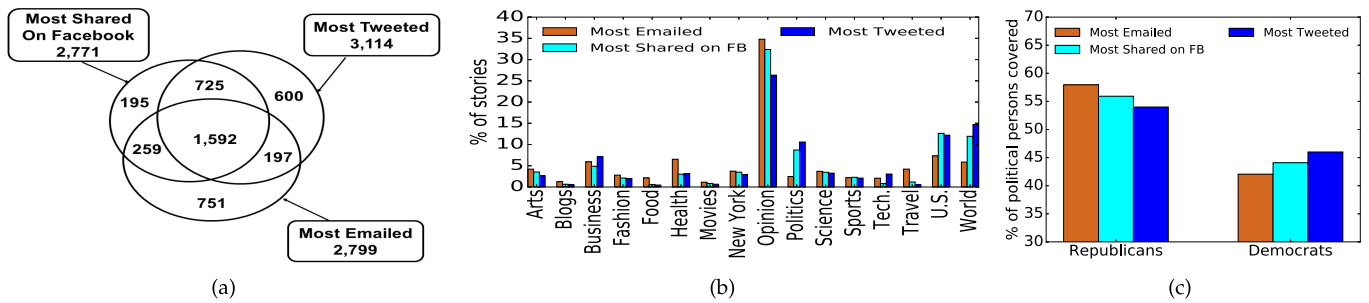


Fig. 2. Comparing audience-selections across different social media—email, Facebook, and Twitter—according to (a) overlap of individual stories, (b) their sectional coverage, and (c) coverage of political persons.

(e.g., U.S. presidential election) shows that audience curation can also have undesirable consequences (such as the spread of false news or extreme opinions) [25]. This debate points to the need to quantify the extent to which the audience and the editors differ in determining which news are important; this section attempts to address that need.

### B. Comparing Audience Selections Across Different Social Sharing Channels

In this section, we focus on comparing the news stories that are selected by the audience for sharing over three media platforms: 1) most emailed stories; 2) stories most shared on Facebook; and 3) stories that are most tweeted.

1) *Difference in Coverage Of Individual Stories:* Fig. 2(a) shows a Venn diagram that represents the overlap in most emailed, most tweeted, and most shared stories on Facebook. We observe that there are significant fractions of every set of stories, which are *not* covered by the other sets. For instance, the set of stories most shared by audience on Facebook would not cover 26% ( $600 + 197 = 797$  out of 3114) of the stories most shared on Twitter; on the other hand, stories most tweeted would not cover 36% ( $751 + 259 = 1010$  out of 2799) of the stories that are most emailed.

In a seminal work, Goffman [26] argued that people consider specific social contexts while presenting themselves in different interactions, where they want to control their impressions in others' eyes. Carrying forward this argument, Hogan [27] postulated that the use of social media is an enduring *exhibition* of one's self-disclosed online identity. We similarly observe that across multiple social media, people tend to maintain different identities by sharing different types of stories.

2) *Difference in Sectional Coverage:* Next, Fig. 2(b) shows the comparison of the sectional coverage of audience-selected stories on different media. To compare the three media—email, Facebook, and Twitter—we run three two-sample t-tests for every section to find whether there are statistically significant differences in the coverage of the sections. For the sections shown in Fig. 2(b), the differences in coverage are statistically significant (similar to the earlier case, the significance level is 0.025, after applying the Bonferroni Correction [15]).

We observe that stories belonging to certain sections like Travel, Opinion, and Food as well as Health stories are

shared more via email than via social media like Facebook or Twitter. On the other hand, news stories on Politics, U.S., or World are more shared on social media than over email. We also see differences in sectional coverage of stories shared over different social media (Facebook and Twitter). World, Technology, and Business stories are more tweeted, whereas Arts, Opinion, and national stories (i.e., on U.S.) are more shared on Facebook.

A long line of works in the communication theory has focused on the notion of an *imagined audience*, which is a person's cognitive idea of the intended recipient of a communication [28]. Researchers have found that the imagined audience is as effective as the actual audience in determining one's communication behavior [29]. When people share news over social media, without being able to see the actual audience (in contrast to what happens in face-to-face interactions), they create an imagined audience for their intended recipient. With different neighborhoods in different social networks, the imagined audience does not remain stable across different media, and hence, people tend to share differently considering their imagined audience for that particular medium. For instance, email (mostly one-to-one communication) is being used by the audience for sharing more personal stories (on Fashion, Food, or Health). On the other hand, Facebook (mostly conversation among reciprocal friends) is being used for sharing of national or local news. Finally, Twitter (one-to-many followers communication) is used for sharing Business, Politics, or World stories.

3) *Difference in Topical Coverage:* To compare the topical coverage of stories selected by the audience to share over different media, we find the topics covered more prominently in one selection by using Fisher's exact test as described earlier. Table V gives some example topics that are covered prominently either for sharing on Twitter or on Facebook. Similarly, Table VI gives the example topics covered prominently either for sharing on Twitter or over email.

We observe in Tables V and VI that audience prominently covers mostly political topics to share on Twitter, whereas the topics prominently covered by audience to share on Facebook are historical events or lifestyle topics. On the other hand, most emailed stories mostly cover niche topics such as "Travel and Vacations" or "Cooking and Cookbooks."

We also see the difference in the geographical coverage. Most of the locations covered on Twitter are *international*,

TABLE V  
TOPICS PROMINENTLY COVERED BY THE AUDIENCE TO SHARE ON TWITTER OR ON FACEBOOK

Topic type	Example topics prominently covered by the audience to share on Twitter, but not on Facebook
Subject	Courts and the Judiciary, Civil Rights and Liberties, Political Action Committees, Mobile Applications, International Trade and World Market, Banking and Financial Institutions, Work-Life Balance
Location	Ukraine, Yemen, North Korea, Scotland, Croatia, Silicon Valley, Las Vegas, Gaza Strip, Damascus, Ramadi, Ankara
Person	Abdel Fattah el Sisi, David Cameron, Shinzo Abe, Raul Castro, Aung Suu Kyi, Bill Gates, Elon Musk
Organization	Al Jazeera, United Nations High Commissioner for Refugees, Iraqi Army, Taliban, Free Syrian Army, Amnesty International, Doctors Without Borders, Amazon.com, Airbnb, Netflix, Playboy Magazine
Topic type	Example topics prominently covered by the audience to share on Facebook, but not on Twitter
Subject	Israeli Settlements, Holocaust and the Nazi Era, World War II, Land Use Policies, Social Security (US), Minimum Wage, Religion and Belief, Jews and Judaism, Pregnancy and Childbirth, Parenting
Location	Ohio, Illinois, North Carolina, South China Sea, Vietnam, Manhattan (NYC), Brooklyn (NYC), Boston
Person	Bobby Jindal, Annise D. Parker, Oliver Sacks, Julian Bond, Yogi Berra, Harrison Ford, Roberta Vinci
Organization	Select Committee on Benghazi, New York Mets, New York Yankees, Chicago Cubs, Cincinnati Reds, Harvard Medical School, Stanford University, American Museum of Natural History

TABLE VI  
TOPICS PROMINENTLY COVERED BY THE AUDIENCE TO SHARE ON TWITTER OR ON EMAIL

Topic type	Example topics prominently covered by audience to share on Twitter, but not on email
Subject	United States International Relations, Middle East and Africa Migrant Crisis, Police Brutality, Campaign Finance, Layoffs and Job Reductions, Federal Taxes (US)
Location	Iran, Afghanistan, Egypt, Pakistan, Venezuela, West Bank, Raqqa (Syria), Brussels, Guatemala
Person	Bashar al-Assad, Xi Jinping, Hassan Rouhani, Recep Erdogan, Bill Cosby, Huma Abedin
Organization	Federal Bureau of Investigation, Justice Department, Afghan National Security Forces, Federal Election Commission, National Security Agency, European Commission, Hezbollah, Palestinian Authority
Topic type	Example topics prominently covered by audience to share on email, but not on Twitter
Subject	Travel and Vacations, Retirement, Books and Literature, Writing and Writers, Elderly, Cooking and Cookbooks, Restaurants, Medicine and Health, Death and Dying, Anxiety and Stress, Theater
Location	Florida, Western States (US), East Coast (US), Italy, Japan, East Village (Manhattan, NY), Boston (Mass), Jackson Heights (Queens, NY), Times Square and 42nd Street (Manhattan, NY), Rhode Island
Person	Sarah Palin, Roger Federer, Serena Williams, Daniel Craig, Lena Dunham, Pablo Picasso
Organization	Yale University, New England Journal of Medicine, New York Public Library, American Cancer Society, Health and Human Services Department

whereas locations covered on Facebook and email are more *national and local* (i.e., within U.S.). Similarly, people covered on Twitter are mostly business tycoons or heads of different governments; whereas, the persons covered on Facebook or email are U.S. politicians, movie actors, or sports personalities. The organizations covered on Twitter are political actors or business entities; whereas, organizations covered on Facebook or email are Sports teams and educational entities.

To check the coverage of political persons in news stories shared over different media, we plot in Fig. 2(c), the percentage of republicans and democrats being covered. Across all three media, we observe republican people tend to be covered more than democrats, with the highest share of republicans in stories shared over email and least in stories shared on Twitter.

4) *Summary*: We observe that more political and international topics are shared on Twitter as against more lifestyle and local topics getting shared on email and Facebook. Thus, the news audience are selecting stories depending on different recipients on different platforms (close friends on email, acquaintances on Facebook, and followers on Twitter). This is in line with the works in the communication theory that postulates that the difference in sharing behavior of an individual across different channels stems from the attempt

to control his/her impressions in the eyes of his/her imagined audience [26], [28].

#### IV. GENERALIZABILITY

Up until now, we have analyzed the data collected from NYTimes during the run-up to the 2016 U.S. presidential election. To check whether the observations generalize to other scenarios, in this section, we investigate NYTimes stories from a different time period as well as the editor and audience selection of news stories of a different media website.

##### A. Analyzing NYTimes Stories at a Different Period

We continued collecting data from NYTimes beyond the time period reported in Section II-A. In this section, we look into data from November to December 2017, where we analyze the top stories, most viewed stories, and stories most shared over different channels. Similar to the results reported in Section III, we observe that there are significant differences between the stories selected by the editors and the audience, as well as between stories shared by the audience over different social media channels.

Fig. 3(a) shows the sectional coverage of editor-selected top stories versus audience-selected most viewed and most



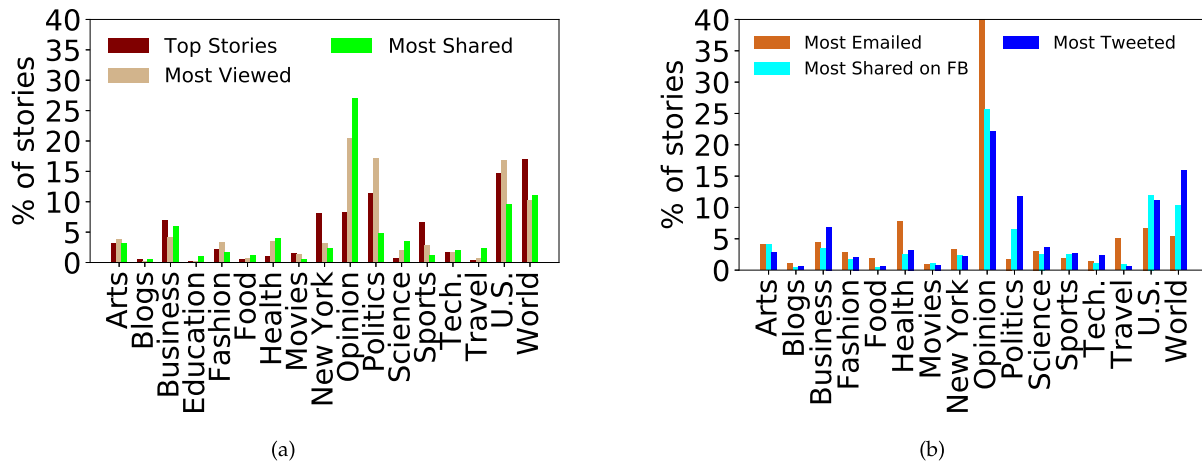


Fig. 3. Comparing the sectional coverage of (a) editor-selected versus audience-selected NYTimes stories and (b) audience-selected NYTimes stories for sharing over Facebook, email, and Twitter during a different time period: November–December 2017.

shared stories. Fig. 3(b) shows the sectional coverage of audience-selected NYTimes stories for sharing over Facebook, email, and Twitter. We can observe from the figures that although the exact coverage of different sections are quantitatively different from the corresponding results in Section III [i.e., Figs. 1(b) and 2(b)], the trends are qualitatively similar to what we observed earlier. For example, political stories were more shared on Twitter compared to Facebook even during the last two months of 2017. Thus, we can conclude that the results reported in Section III are not the manifestation of a particular event (e.g., 2016 U.S. presidential election), rather it reflects the changed news ecosystem and can generalize to other time periods as well.

### B. Extending Observations to Another Media

Next, we investigate the editor selection and the audience selection of news stories at another popular news media—The Guardian ([theguardian.com](http://theguardian.com)), with a specific focus on its U.K. edition. The Guardian homepage features both the editor selected “Headline” stories (at the top of the homepage) and the audience selected “Most Viewed” stories (toward the bottom of the page). Although, similar to NYTimes, The Guardian also provides social media sharing buttons (e.g., over Facebook, Twitter, or Email) against every news story, the lists of most shared stories on these media are not publicly available. Hence, we restrict our focus on comparing editorial selection (in headline stories) and selective exposure by audience (as manifested in most viewed stories).

1) *Data Set Collected*: The Internet Archive,<sup>5</sup> a nonprofit organization for conservation of Internet information, maintains timely snapshots of different sites across the Web. Every day, it captures on average 25 snapshots for The Guardian homepage at different times of the day. We crawled the snapshots of The Guardian homepage captured by Internet Archive between July 2015 and February 2016, and the same 8-month period covered in the NYTimes data set studied

in Section III. Then, we extracted the list of headline stories and most viewed stories from the collected snapshots.

In addition, using The Guardian Developer API,<sup>6</sup> we collected all the articles published during the 8-month period including the content and metadata (e.g., the author, section, keywords, etc.) associated with every story. Overall, we collected information for 11 587 unique news stories out of which 8473 stories were part of the list of headline stories or most viewed stories at least once during this 8-month period.

2) *Difference in Individual News Stories*: We compute the overlap and difference in the set of headline stories and most viewed stories and observe that similar to NYTimes, The Guardian audience choose a large fraction (53.4%) of stories that are not promoted by the expert editors [shown in Fig. 4(a)]. Thus, we can conclude that the reliance on expert editorial control is decreasing regardless of the media site being studied.

3) *Difference in Sectional Coverage*: Next, we compare the sectional coverage of headline and most viewed stories, and Fig. 4(b) shows the result. We find statistically significant differences ( $p < 0.05$ ) in the coverage of all sections in choices of the editors and the news audience. We observe in Fig. 4(b) that The Guardian stories belonging to news sections like Business, Politics, U.K., or World are less selected by the audience (compared to their relative coverage in editor selections).

On the other hand, stories belonging to sections like Film, Football, Lifestyle, or TV, as well as “Opinion” pieces are selected by the audience way more than their share in the editor-selected headline stories. Recall that we observed similar trends while comparing editorial and audience selection of NYTimes stories.

4) *Difference in Topical Coverage*: Similar to NYTimes, The Guardian also provides a set of topics (tags) for every news story [30]. We explore the topics covered more prominently in either editor or audience selection in Table VII. We find that The Guardian editors prominently cover mostly

<sup>5</sup><https://archive.org>

<sup>6</sup><https://open-platform.theguardian.com>



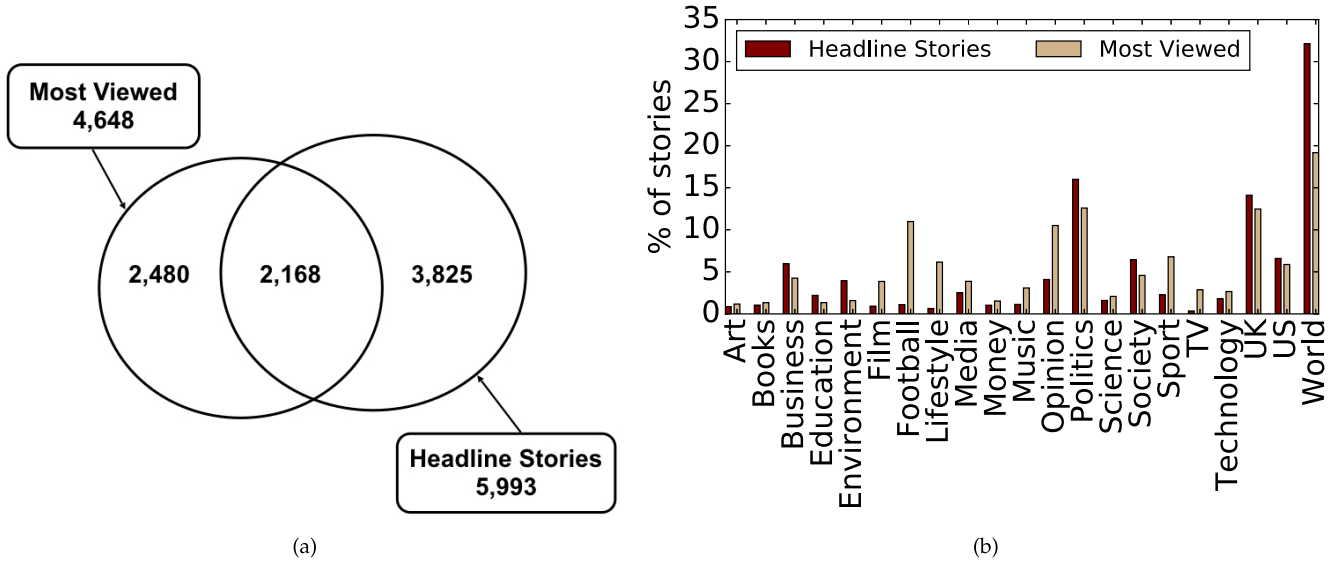


Fig. 4. Comparing editor-selected headline stories versus audience-selected most viewed stories on The Guardian, according to (a) overlap of individual stories and (b) sectional coverage.

TABLE VII  
TOPICS PROMINENTLY COVERED BY EITHER THE EDITORS OR THE AUDIENCE OF THE GUARDIAN

Example topics prominently covered by the editors, but not by the audience
Refugees, Climate change, Military, Defence policy, United Nations, Tax and spending, Eurozone crisis, Nuclear weapons, UK criminal justice, House of Lords, European commission, Syria, Human rights, South and Central Asia, Stock markets, FIFA, Charities, Air transport
Example topics prominently covered by the audience, but not by the editors
Race issues, LGBT rights, Feminism, Rape and sexual assault, Manchester United, Liverpool, Chelsea, Manchester City, Food & drink, Relationships, Health & wellbeing, Australia sport, Parents and parenting, Celebrity, Rugby World Cup, Science fiction and fantasy, Star Wars

political or economy related topics and only a few niche topics such as “FIFA” or “Charities.” On the other hand, the audience prominently cover more human interest or lifestyle topics and less political topics.

5) *How Different is Guardian From NYTimes?:* In this section, we attempted to investigate whether our observations on the editorial and audience selections of NYTimes stories generalize to other media organizations, and we chose The Guardian for that purpose. Next, we compare the coverage of stories published by NYTimes and The Guardian to check whether the media houses publish similar or different type of stories.

Fig. 5 shows the sectional coverage of stories published by NYTimes and The Guardian. We can observe in Fig. 5 that the media sites have a disparate distribution of stories published on different sections. For example, stories on Entertainment, Fashion, and World are more published by The Guardian compared to NYTimes. On the other hand, NYTimes tends to cover more national issues, business, and sports news than The Guardian.

Table VIII gives the topics covered more prominently by either of the media organizations. As NYTimes and The Guardian focus on news from two different countries, their topical coverage differs significantly. For example, NYTimes covers U.S.-related topics (e.g., United States International Relations and Federal Budget) more prominently

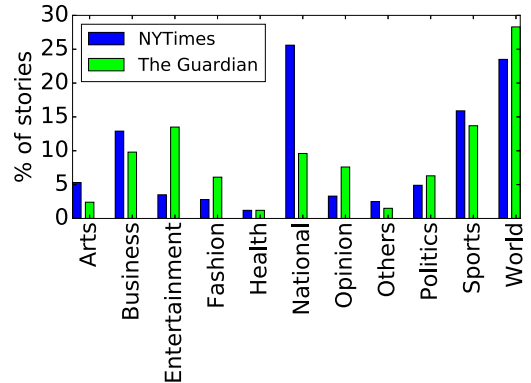


Fig. 5. Comparing sectional coverage of stories published in NYTimes and The Guardian. Here, for better comparison, we have manually merged several sections in both sites into common themes.

than The Guardian. On the other hand, The Guardian covers U.K.-related topics (e.g., House of Lords and National Health Service) more prominently. Thus, we can conclude that the media organizations produce stories differently from one another focusing on different sets of issues.

6) *Summary:* We repeated the comparison between the editor and audience selection on a different media site from a different country than NYTimes. Yet, we observed almost identical trends across these two media sites. The results

TABLE VIII  
TOPICS PROMINENTLY COVERED IN NEWS STORIES PUBLISHED BY EITHER NYTIMES OR THE GUARDIAN

<b>Example topics prominently covered by NYTimes, but not by The Guardian</b>
United States International Relations, Police Brutality, Misconduct and Shootings, Illegal Immigration, Federal Budget, Demonstrations, Protests and Riots, Income Inequality, Stocks and Bonds, Computers and the Internet
<b>Example topics prominently covered by The Guardian, but not by NYTimes</b>
House of Lords, European Commission, UK Criminal Justice, Metropolitan Police, NHS, Manchester United, Liverpool, Chelsea, Eurozone Crisis, Northern Ireland, Pop and Rock, England Cricket Team

further demonstrate that in the current media landscape, the news audience employs far more control over the type of news they want to read. Although NYTimes and The Guardian may have very different editorial selection criteria, both of them highlight more *hard news* (e.g., political and finance news) in their selections. However, we observed that the audience in both media sites tend to favor *softer* lifestyle or other human-interest stories.

## V. RELATED WORK

In this section, we briefly discuss the related research efforts.

### A. Journalistic Gatekeeping

Gatekeeping has been described as the process by which journalists and editors select potential news stories for circulation [1]. Considering the importance of editorial selections in setting the news discourse in a society, media researchers have extensively analyzed the selected news articles on the issues of *bias* as well as *fairness and accuracy* of the presented facts [31]–[33]. Moreover, media watchdog groups like FAIR (fair.org) monitor the media organizations for any bias or misinformation in news stories. However, in the online media landscape, as editorial calculations do not remain the sole domain of the expert editors, there is a need to analyze the selections of the other stakeholder—the news audience.

### B. Selective Exposure

Selective exposure can occur when the selected messages diverge from the composition of all accessible messages [34]. It enables people to defend their attitudes, beliefs, and behaviors by avoiding information likely to challenge them, a phenomenon known as *congeniality or confirmation bias* [35]. Prior works in media studies have investigated selective exposure in connection with print media [36]. However, the ease of selectivity in online news consumption (and the popularity of *personalized recommendations* [37], where new content is recommended based on user’s past behavior) has increased the possibility of people getting trapped in echo chambers or “filter bubbles” [38]. In this paper, we analyze the selective exposure of the news audience as a group, which may differ from individual selections.

### C. Exploring the News Gap

Traditionally, prior works in media studies (such as by Hart [39]) have compared the topical coverage of different

printed newspapers, thereby comparing different editorial selections. With online news consumption becoming mainstream, researchers observed differences between online and offline consumption. For instance, Althaus and Tewksbury [2] found that the editorial selections on print media controlled the agenda-setting power with the audience taking their cues; whereas, online audience-selected topics more freely across news sections. More recently, Boczkowski and Mitchelstein [7] explored the “news gap” between editor selection and audience consumption in multiple news media sites. They found that while the editors select public affairs (e.g., national, international, and business stories) to be most important, audience consumption moves toward nonpublic affairs content (e.g., sports, crime, and entertainment). We extend these prior attempts to not only uncover the gap between editor and audience news preferences but also the preferences of audience for sharing across different media [40].

### D. Social News Sharing and Consumption

With the increasing popularity of social media, there has been a close interplay between news media and social media sites. Recognizing the incoming traffic from social media, news websites have also provided tools such as news sharing buttons, which facilitate and simplify social news sharing by the audience. As an effect, a recent survey by Reuters Institute found that 44% of respondents regularly use social media Facebook to consume and discuss news stories [41]. Multiple research works have focused on news coverage in social media. Broersma and Graham [42] studied the British election coverage on Twitter. Bakshy *et al.* [43] analyzed the amount of politically cross-cutting contents in Facebook. Ribeiro *et al.* [44] proposed a mechanism to utilize the intended audience of contents to measure the bias of social media news outlets.

There have also been studies exploring news coverage across social media sites. Bright [45] studied how different topics are shared across social media site and found that economics and finance news is shared more on LinkedIn, but less on Facebook; whereas, human interest news is less popular on LinkedIn than Facebook. Bastos [8] observed that audience on Facebook share more stories on arts, but stories shared on Twitter contain a higher percentage of economy, technology, and national news.

In all of these prior works, the focus has been either on editorial selection versus audience consumption or the audience sharing in social networks. To the best of our knowledge,

in this paper, we perform the first study that combines all three together—editorial selection, selective audience consumption, and network gatekeeping while sharing news.

## VI. CONCLUSION

In this paper, we tried to understand the interactions among three different dynamics associated with the consumption and propagation of news—editors’ selection, readers’ consumption, and readers’ sharing pattern. We found that the interactions between these three dynamics have made the news distribution pattern fairly complex—a huge change from the traditional editor-driven strategy.

We can further contextualize this change by drawing from the seminal work on the “Power of Identity” by Castells [46], where he theorized how individuals and groups are interrelated in different layers creating different identities within the broader framework of *networked society*. Castells [46] differentiated between three types of identities.

- 1) *Legitimizing Identity*: It is introduced by the dominant social institutions to extend and rationalize their domination over social actors.
- 2) *Resistance Identity*: It is generated by the actors being excluded by the logic of domination and thus building resistance based on principles different from emanating from the institutions.
- 3) *Project Identity*: These are proactive actions to change the status quo rather than merely surviving in the opposition of the dominant institutions.

In our news media context, we clearly see these three identities taking shape. The media organizations, from their implicit goal of shaping public opinion, try to create a legitimizing identity of their audience by actively highlighting the stories which they deem to be important. However, the news media audience attempt to create a resistance identity by taking a different path in consuming stories that are at odds with the media organizations want. Finally, due to the power of social sharing, the audience actively create a project identity where they decide to draw an alternate landscape based on the stories they deem fit for the broader society to consume. Interestingly, the project identity takes different shapes depending on the confluence between online and offline identities, resulting in differences in news selected for sharing via different communication channels.

There may be several long-term implications of the tension between these different identities. For example, the significant differences among the stories the audience select for consumption as against sharing to their peers also may create “news gap” for a user consuming news primarily through social media. This may lead to the rise of social media only news outlets (initial signs are already seen), where there are concerns about the bias and authenticity of news propagated by such outlets [47], [48].

Anecdotal evidence suggests that the media organizations are trying to adapt to the changed landscape by: 1) optimizing their story-promotion and 2) planning social media strategies to meet the crowd demand. However, such approaches have the risk of lowering the quality of news (e.g., proliferation of *clickbaits* [49] and increased *tabloidization* of news [50], [51]).

Moreover, as there are differences in crowd choices across different media platforms, a single strategy may not work and the media organizations need to tailor their promotions on different platforms. We believe that we are yet to realize all angles associated with the changed media landscape, and works such as ours would be important steps toward realizing the depth of the problem and finding a holistic solution considering all associated aspects.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers whose suggestions helped to improve this paper.

## REFERENCES

- [1] P. J. Shoemaker and T. Vos, *Gatekeeping Theory*. Abingdon, U.K.: Routledge, 2009.
- [2] S. L. Althaus and D. Tewksbury, “Agenda setting and the ‘new’ news: Patterns of issue importance among readers of the paper and online versions of the New York times,” *Commun. Res.*, vol. 29, no. 2, pp. 180–207, 2002.
- [3] D. O. Sears and J. L. Freedman, “Selective exposure to information: A critical review,” *Public Opinion Quart.*, vol. 31, no. 2, pp. 194–213, 1967.
- [4] N. J. Stroud, “Polarization and partisan selective exposure,” *J. Commun.*, vol. 60, no. 3, pp. 556–576, 2010.
- [5] A. Mitchell, J. Gottfried, J. Kiley, and K. E. Matsa. (Oct. 2014). *Social Media, Political News and Ideology*. [Online]. Available: <https://journalism.org/2014/10/21/section-2-social-media-political-news-and-ideology>
- [6] K. Barzilai-Nahon, “Toward a theory of network gatekeeping: A framework for exploring information control,” *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 9, pp. 1493–1512, 2008.
- [7] P. J. Boczkowski and E. Mithelstein, *The News Gap—When the Information Preferences of the Media and the Public Diverge*. Cambridge, MA, USA: MIT Press, 2013.
- [8] M. T. Bastos, “Shares, pins, and tweets: News readership from daily papers to social media,” *J. Stud.*, vol. 16, no. 3, pp. 305–325, 2015.
- [9] N. Y. Times. (2016). *Article Search API*. [Online]. Available: [developer.nytimes.com/article\\_search\\_v2.json](https://developer.nytimes.com/article_search_v2.json)
- [10] N. Carlson. (2009). *NYT.com Front Page Editors Ignore Reader Clicks*. [Online]. Available: [goo.gl/i1NTq](https://goo.gl/i1NTq)
- [11] N. Y. Times. (2016). *The Top Stories API*. [Online]. Available: [developer.nytimes.com/top\\_stories\\_v2.json](https://developer.nytimes.com/top_stories_v2.json)
- [12] (2016). *The Most Popular API*. [Online]. Available: [https://developer.nytimes.com/most\\_popular\\_api\\_v2.json](https://developer.nytimes.com/most_popular_api_v2.json)
- [13] G. D. Ruxton, “The unequal variance t-test is an underused alternative to Student’s t-test and the Mann–Whitney *U* test,” *Behav. Ecol.*, vol. 17, no. 4, pp. 688–690, 2006.
- [14] B. L. Welch, “The generalization of ‘Student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, pp. 28–35, Jan. 1947.
- [15] C. E. Bonferroni, “Il calcolo delle assicurazioni su gruppi di teste,” in *Studi in Onore del Professore Salvatore Ortu Carboni*. Rome, Italy, 1935, pp. 13–60.
- [16] J. Harris. (Oct. 2007). *Messing Around With Metadata*. [Online]. Available: <https://open.blogs.nytimes.com/2007/10/23/messing-around-with-metadata>
- [17] R. Routledge, “Fisher’s exact test,” in *Encyclopedia of Biostatistics*. Hoboken, NJ, USA: Wiley, 2005.
- [18] G. J. G. Upton, “Fisher’s exact test,” *J. Roy. Stat. Soc.*, vol. 155, no. 3, pp. 395–402, 1992.
- [19] M. McCombs, *The News and Public Opinion: Media Effects on Civic Life*. Hoboken, NJ, USA: Wiley, 2011.
- [20] K. D. Harber and D. J. Cohen, “The emotional broadcaster theory of social sharing,” *J. Language Social Psychol.*, vol. 24, no. 4, pp. 382–400, 2005.
- [21] P. J. Boczkowski, *News at Work*. Chicago, IL, USA: Chicago Univ., 2010.
- [22] M. T. Stratton, “Uncovering a new guilty pleasure: A qualitative study of the emotions of personal Web usage at work,” *J. Leadership Organizational Stud.*, vol. 17, no. 4, pp. 392–410, 2010.



- [23] M. L. Richins and T. Root-Shaffer, "The role of involvement and opinion leadership in consumer word-of-mouth: An implicit model made explicit," *NA—Advances in Consumer Research*. Provo, UT, USA: Association for Consumer Research, 1988.
- [24] L. Spayd. (2016). *Why Readers See the Times as Liberal*. [Online]. Available: <https://www.nytimes.com/2016/07/24/public-editor/liz-spayd-the-new-york-times-public-editor.html>
- [25] O. Solon. (2016). *Facebook's Failure: Did Fake News and Polarized Politics Get Trump Elected?* [Online]. Available: [theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories](http://theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories)
- [26] E. Goffman, *The Presentation of Self in Everyday Life*. New York, NY, USA: Doubleday, 1978.
- [27] B. Hogan, "The presentation of self in the age of social media: Distinguishing performances and exhibitions online," *Bull. Sci., Technol. Soc.*, vol. 30, no. 6, pp. 377–386, 2010.
- [28] E. Litt, "Knock, Knock. Who's there? The imagined audience," *J. Broadcast. Electron. Media*, vol. 56, no. 3, pp. 330–345, 2012.
- [29] A. J. Fridlund, "Sociality of solitary smiling: Potentiation by an implicit audience," *J. Personality Social Psychol.*, vol. 60, no. 2, pp. 229–240, 1991.
- [30] M. Belam and P. Martin. (Feb. 2011). *Tags are Magic*. [Online]. Available: [theguardian.com/info/series/tags-are-magic](http://theguardian.com/info/series/tags-are-magic)
- [31] D. D'Alessio and M. Allen, "Media bias in presidential elections: A meta-analysis," *J. Commun.*, vol. 50, no. 4, pp. 133–156, 2000.
- [32] T. Groeling, "Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news," *Political Sci.*, vol. 16, no. 1, pp. 129–151, 2013.
- [33] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Can trending news stories create coverage bias? On the impact of high content churn in online news media," in *Proc. Comput. Journalism Symp.*, 2015, pp. 1–5.
- [34] F. Arendt, N. Steindl, and A. Kumpel, "Implicit and explicit attitudes as predictors of gatekeeping, selective exposure, and news sharing: Testing a general model of media-related selection," *J. Commun.*, vol. 66, no. 5, pp. 717–740, 2016.
- [35] W. Hart, D. Albarracín, A. H. Eagly, I. Brechan, M. J. Lindberg, and L. Merrill, "Feeling validated versus being correct: A meta-analysis of selective exposure to information," *Psychol. Bull.*, vol. 135, no. 4, pp. 555–588, 2009.
- [36] P. J. Tichenor, G. A. Donohue, and C. N. Olien, "Mass media flow and differential growth in knowledge," *Public Opinion Quart.*, vol. 34, no. 2, pp. 159–170, 1970.
- [37] A. Chakraborty and N. Ganguly, "Analyzing the news coverage of personalized newspapers," in *Proc. IEEE/ACM ASONAM*, Aug. 2018, pp. 540–543.
- [38] E. Pariser, *The Filter Bubble: What the Internet is Hiding From You*. London, U.K.: Penguin, 2011.
- [39] J. A. Hart, "Foreign news in U. S. and english daily newspapers: A comparison," *J. Mass Commun. Quart.*, vol. 43, no. 3, pp. 443–448, 1966.
- [40] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Dissemination biases of social media channels: On the topical coverage of socially shared news," in *Proc. AAAI ICWSM*, 2016, pp. 559–562.
- [41] N. Newman, D. A. Levy, and R. K. Nielsen, "Digital news report 2016," Reuters Inst. Study Journalism, Oxford, U.K., Tech. Rep., 2016.
- [42] M. Broersma and T. Graham, "Social media as beat: Tweets as a news source during the 2010 British and Dutch elections," *J. Pract.*, vol. 6, no. 3, pp. 403–419, 2012.
- [43] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.
- [44] F. N. Ribeiro *et al.*, "Media bias monitor: Quantifying biases of social media news outlets at large-scale," in *Proc. AAAI ICWSM*, 2018.
- [45] J. Bright, "The social news gap: How news reading and news sharing diverge," *J. Commun.*, vol. 66, no. 3, pp. 343–365, 2016.
- [46] M. Castells, *The Power of Identity: The Information Age: Economy, Society and Culture*. Oxford, U.K.: Blackwell, 1997.
- [47] M. Del Vicario *et al.*, "The spreading of misinformation online," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 3, pp. 554–559, 2016.
- [48] M. Babaei, A. Chakraborty, J. Kulshrestha, E. M. Redmiles, M. Cha, and K. P. Gummadi, "Analyzing biases in perception of truth in news stories and their implications for fact checking," in *Proc. ACM FAT*, 2019, p. 139.
- [49] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop clickbait: Detecting and preventing clickbaits in online news media," in *Proc. IEEE/ACM ASONAM*, Aug. 2016, pp. 9–16.

- [50] A. Chakraborty, R. Sarkar, A. Mrigen, and N. Ganguly, "Tabloids in the era of social media?: Understanding the production and consumption of clickbaits in Twitter," in *Proc. ACM Hum. Comput. Interact.*, vol. 1, Nov. 2017, Art. no. 30.
- [51] J. B. Mackay and E. Bailey, "Sacrificing credibility for sleaze: Mainstream media's use of tabloidization," in *Web Design Development: Concepts, Methodologies, Tools, Appl.* IGI Global, 2016.



**Abhijnan Chakraborty** received the Ph.D. degree from IIT Kharagpur, Kharagpur, India, in 2019.

He is currently a Post-Doctoral Researcher with the Max Planck Institute for Software Systems (MPI-SWS), Saarbrücken, Germany. His current research interests include social computing, information retrieval, and fairness in algorithmic decision making. He has authored several papers in top-tier venues in these areas.

Dr. Chakraborty's research has won the Best Paper Award at ASONAM'16 and the Best Poster Award at ECIR'19.



**Saptarshi Ghosh** received the Ph.D. degree from IIT Kharagpur, Kharagpur, India, in 2013.

He was a Humboldt Post-Doctoral Fellow with the Max Planck Institute for Software Systems, Saarbrücken, Germany. He is currently an Assistant Professor with the Department of Computer Science and Engineering, IIT Kharagpur. His current research interests include social network analysis, data mining, and information retrieval. He has authored or coauthored more than 50 publications in reputed journals and conferences in these domains.



**Niloy Ganguly (SM'19)** received the B.Tech. degree from IIT Kharagpur, Kharagpur, India, in 1992, and the Ph.D. degree from IIST Shibpur, Kolkata, India, in 2004.

He was a Post-Doctoral Fellow with the Dresden University of Technology, Dresden, Germany. In 2005, he joined the Department of Computer Science and Engineering, IIT Kharagpur, where he is currently a Professor. He is also the Fellow of the Indian National Academy of Engineering, Gurugram, India. His current research interests

include network science, social computing, and machine learning.



**Krishna P. Gummadi** received the B.Tech. degree from IIT Madras, Chennai, India, and the Ph.D. degree from the University of Washington, Seattle, WA, USA.

He is currently the Head of the Networked Systems Research Group, Max Planck Institute for Software Systems (MPI-SWS), Saarbrücken, Germany, and a Professor with the University of Saarland, Saarbrücken. His current research interests include understanding and building social computing systems. His current projects focus on enhancing fairness, accountability, and explainability of automated (particularly, data-driven and learning-based) decision-making systems.

Dr. Gummadi was a recipient of numerous awards including the ACM SIGCOMM Test-of-Time Award.