Breaking News and Younger Twitter Users: Comparing Self-Reported Motivations to Online Behavior

Kenneth Joseph
Computer Science and Engineering
University at Buffalo
Buffalo, NY
kjoseph@buffalo.edu

John Wihbey
School of Journalism
Northeastern University
Boston, MA
j.wihbey@northeastern.edu

ABSTRACT

Breaking news represents an important avenue of information about current events, including life-or-death information during natural disasters, but it can also serve as a vector along which misinformation can rapidly spread. In the present work we study factors associated with the sharing of breaking news by young, college-aged students. Using a unique combination of survey and behavioral data, we identify traits associated with a propensity to share breaking news on Twitter. We find that individuals who share more breaking news report high levels of confidence in their ability to differentiate real from fake news and to manage information overload. However, breaking news sharing is not associated with the reported use of traditional fact-checking strategies (e.g., finding other sources for the same information) before sharing. Thus, our data are consistent with the idea that breaking news sharers tend to rely, at least in part, on confidence in their own understanding of the news when determining what breaking news to share. We contextualize these findings by studying patterns of general news sharing and non-news sharing, as well as studying connections between party affiliation and factors associated with the sharing of content.

CCS CONCEPTS

• **Networks → Social media networks**; **Online social networks**; • **Human-centered computing → Social content sharing**; **Empirical studies in collaborative and social computing**.

KEYWORDS

breaking news, Twitter, computational social science, journalism, students

ACM Reference Format:
Kenneth Joseph and John Wihbey. 2019. Breaking News and Younger Twitter Users: Comparing Self-Reported Motivations to Online Behavior. In *International Conference on Social Media and Society (SMSociety '19), July 19–21, 2019, Toronto, ON, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3328529.3328548

SMSociety '19, July 19–21, 2019, Toronto, ON, Canada

ACM ISBN 978-1-4503-6651-9/19/07.
https://doi.org/10.1145/3328529.3328548

1 INTRODUCTION

Social media platforms and online sources account for an increasingly large portion of news access and consumption, and young, college-aged persons in particular are more likely to access news online than through traditional news media pathways [12].

With social media platforms such as Twitter and Facebook, the news served up to individuals through streams and feeds is determined by a combination of active individual choice— selection of whom to follow or intentional network creation— and by algorithms that draw upon a vast array of social signals that help determine the relevance of content and rank it [34]. This means that the news presented to individuals online, while often originating from mainstream media outlets [13], may first filter through myriad paths of intentional network creation and algorithmic curation. Individuals who participate actively and share content widely within these platforms may influence the processes and news outputs of such "social algorithms" [18].

The gatekeeping function of traditional news media is therefore being partially supplanted and substantially complemented by individuals and groups, who through their activity and influence on social media can help to shape what users see and consume. These shifts demand new theories, for instance network gatekeeping theory, which attempt to account for these complex dynamics emerging with respect to technology and media, as the dimensions of information control in society change and evolve in ways that prior exposure-related models cannot capture [1, 23].

Research in the network gatekeeping tradition can be seen as a piece of a broader argument that understanding citizens' media exposure in the web era involves mapping a wide number of variables within a complex "curated flow" model. Exposure increasingly comes from social contacts and individual media users [34]. Such work emphasizes in particular the existence of a new generation of highly active media users, operating as influencers and gatekeepers, who are increasingly shaping media exposure and the public sphere [2, 5]. These local opinion leaders can exert great influence over what gets seen and heard by citizens, and as non-traditional actors can influence information flows and framings around important civic and political events [10].

The present work provides unique perspective on an important subpopulation of these local, non-traditional gatekeepers and their behavior around an important kind of news content. Specifically, we consider who among young persons of college age in the United States may serve as a kind of new gatekeeper for breaking news. We ask:

> *Who among young persons shares breaking news on Twitter, why do they do it, and how does the sharing of breaking news compare to the sharing of news in general, and non-news content?*

We study breaking news because a) it often involves consequential issues and events that impact human lives (as opposed to softer news, feature or lifestyle news, or entertainment/sports coverage) and may have implications for civic health and well-being; and b) it typically involves rapid decision-making on the part of users about whether or not to share, thus serving as a good test of information literacy and judgment. Breaking news might generally be defined as coverage of novel and unfolding "events involving top leaders, major issues, or significant disruptions in the routines of daily life, such as an earthquake or airline disaster" [26].

The consequences of misinformation being rapidly shared can be substantial and, in certain cases, irreparable, damage. Sharing of misinformation on social media has led, for example, to persons being wrongly suspected of criminal activity and even terrorism [25]. The phenomenon is also of increasing concern in diverse societies where ethnic tensions are easily inflamed [31]. The research community continues to explore how misinformation around unfolding crises, both natural and human-induced, should be addressed or combatted [9].

Further, it should be said that our definition of "fake news" in this paper follows [13] and [19] wherein "fakeness" is attributed at the level of the publisher, not the story. Fake news outlets have the "trappings of legitimately produced news" but lack the ethical-editorial norms that characterize traditional journalism institutions, such as fairness, objectivity, and independence from external influence in exercising news judgments. Of course, stories that could qualify as fake news may fall into different analytical categories, such as misinformation or propaganda. Researchers have tended to note these distinctions even as they study the general phenomenon [13].

We choose to study young persons because they access more news on social media, including breaking news, than older cohorts [12], and thus serve as a potentially important subset of non-traditional gatekeepers. Finally, we focus on social media, and Twitter specifically, as it is the place where many young people first receive breaking news [33]. The research literature on news literacy levels continues to grow, and recent work has called into question whether older news consumers are more news literate than younger consumers [11, 13, 15]. How exactly to measure news literacy among younger consumers continues to be debated, as different curricular models are explored in high school and undergraduate education [21, 22].

Understanding the dynamics of information sharing and its causes requires granular data both on user behavioral patterns and motivations. To address this, we draw on a unique research design that allows access, through survey data, to the self-reported factors and motivations of young users on their news sharing behavior (N=3,194). For a subset of these users (N=789), we also have access through the Twitter API to their actual, observable behavior online. The young persons we study are mostly ages 18-24 making them part of the late Millennial and Generation Z (GenZ) cohorts (born in mostly in the mid-1980s and after). They are drawn from a diverse set of colleges and universities across the United States. We examine nearly 1.5 million tweets shared by these individuals, focusing on the roughly 110,000 posts in which links were shared.

To address our research question, we first perform a factor analysis on all survey respondents to identify latent factors that might be related to the sharing of breaking news. For individuals for whom we have Twitter data, we then run a negative binomial regression to find associations between latent factors and the sharing of breaking news. To contextualize our findings for breaking news sharing, we also run similar regressions for news sharing in general, as well as non-news sharing. As described below, we use a combination of manual and automated methods to classify breaking news and news in general.

Our main findings are three-fold:

- Students who claim to find it easier to differentiate real from fake news, and who say they do not feel particularly overwhelmed despite the current, high-velocity news environment, share more breaking news and news in general. However, there is no association between sharing (breaking) news and using active techniques to evaluate information veracity (e.g., evaluating the source of content). Students who share more (breaking) news believe they can differentiate real news from fake news, but do not necessarily claim to take active means to differentiate the two.
- Students who share more news in general, however, do seem to be more conscious of the source of content and generally trust news media more. Even so, these same individuals are not more likely to share breaking news (or to share non-news content). Students who share more non-news content appear to be less concerned about news validity.
- While party affiliation is not a significant predictor of sharing of any kind, the latent factors we identify are distinct across party lines. We therefore find that a students' stated party affiliation is associated with particular latent factors that, in turn, predict sharing behavior.

Given these findings, we discuss the implications for how we might understand the evolving nature of younger active sharers, who may operate as gatekeepers and influencers, and how their practices may help curate the news seen by others in their networks.

2 RELATED WORKS

On social media, certain persons are frequently more likely to be active and to broadcast items, making them potentially influential with peers, particularly with respect to how these peers perceive and trust news [35]. Considerable literature exists studying both the motivations behind this sharing behavior, as well as the content of those shares [17, 20, 32]. In a recent review article, Kümpel et al. [17] identified a variety of factors in the literature that have been associated with news sharing behavior, including higher levels of popularity and activity, and people who also share more in general. Our study aims to complement the existing body of work on sharing behaviors of news content by focusing on 1) differentiating highly active sharers from their peers, 2) young persons of college age and 3) a particularly consequential category of information: breaking news.

Breaking news is important because it may relevant to matters of health and safety, and salient issues of public concern. We focus in particular on Twitter, which has proven a compelling platform for breaking news [27], and has, as mentioned, become a central source of breaking news for young persons [33]. This preference for breaking news on Twitter is due in part to the rapid speed at which information can spread on the platform. The rapid spread of breaking news can be both a boon and a curse. With respect to the former, the rapid spread of breaking news on Twitter during environmental disasters was a critical source of information for those on the ground [30]. With respect to the latter, the very nature of breaking news means that the whole story is not often known, and thus breaking news has been implicated in rumor cascades and the propagation of misinformation [37].

Consequently, the study of breaking news overlaps with the study of why people share, but also with research on the sharing of misinformation on Twitter [13], as well as how perceptions of credibility on Twitter impact sharing behavior [24, 29]. This work has found that factors like education, age, and partisanship can impact perceptions and/or sharing of low credibility content. We complement these efforts by looking at breaking news more generally, and by combining survey and behavioral data to do so.

There have been several other recent efforts to combine survey data with social media data (for a recent, if somewhat selective, review of such work, see [14]). Particularly relevant to our research is the work of Wells and Thorson [36], who combine survey and behavioral data to explore exposure to political shares on Facebook. Additionally, Guess et al. [14] find that survey and behavioral data on sharing are correlated, but that survey data typically underestimates sharing behavior for highly active individuals. Such research underscores the benefit of our methodological approach of combining survey data to understand motivations with behavioral data that more accurately represents user activity than self-reported responses.

3 DATA AND METHODS

In this section, we first give a high-level overview of the survey and Twitter data used to answer our research questions. We then provide more detail on how we identify breaking news shares, general news shares, and non-news shares on Twitter and how we identified latent factors of respondents from the survey data.

3.1 Overview of Survey and Twitter Data

Survey data used here were originally collected by Project Information Literacy[1], a non-profit research group that involves academics from various universities focused on understanding information seeking and sharing behaviors of today's youth. The survey, which the authors of this paper were involved in constructing, focuses broadly on understanding how students engaged with the news, and was carried out in mid 2018. Full details on the survey, in particular details on the sampling methodology and resulting sample, are available at https://www.projectinfolit.org/news_study.html.

The survey included twenty questions, spanning from demographics to questions about news sharing behaviors. Many of these questions had several subparts. One of the questions asked was "If you are a Twitter user, will you share your Twitter username/handle with us?" It was made clear that students could opt out of answering this question, and that their data would be used to study behavioral patterns associated with news sharing and consumption. The survey, as well as these additional analyses on Twitter data, were approved by the Institutional Review Board at Wellesley College.

In total, the survey was completed by 5,844 college students from 11 different universities, colleges and community colleges around the United States and represent, to our knowledge, the largest study of student engagement with news to date. Of these 5,844 respondents, 3,194 provided complete answers to the survey questions studied in the present work, and 890 students responded to the Twitter handle question. Of these responses, 851 could then be linked to a Twitter account. Of these 851, we were able to collect data for 789 of these accounts; the other 61 were protected at all three collection periods. This final set of 789 non-protected accounts represents 13.5% of all students, and 30.4% of respondents who said they used Twitter.

We performed data collection for these Twitter accounts at three different points: April, May and October of 2018. During each collection period, we collected up to their last 3,200 tweets. For the purposes of the present work, we aggregated all data across these three collection periods. We then extracted any tweets containing URLs to websites outside of Twitter, expanding all links in the process. We collected 1,484 tweets from the median college student. In total, we collected 1,425,819 tweets, from which we extract a total of 109,254 URL shares.

3.2 Identifying Breaking News, General News, and Non-News Tweets

We took a high-precision approach to identifying breaking news shares on Twitter, relying on a combination of relatively strict keyword filtering, followed by manual labeling. From all tweets described above, we first selected any tweet where the text of the tweet contained the word "breaking", "developing", "trending", or "exclusive". This resulted in a set of 1,508 tweets, around 1.4% of all tweets containing links to external URLs. For each unique combination of tweet text and linked URL (n=1161, with a subset of these being shared multiple times), one of the authors, an expert on the study of breaking news, manually determined whether or not it was a share of breaking news or not. Of all the tweets containing these keywords, 89.8% of them (1,043) were determined to be shares of breaking news.

To identify all shares as either "news" or "not news", we first expanded URLs and domain names from potential link shorteners, and then classified the domain name for each URL as either "news domains", or "not news domains". To perform this classification, we relied on both manual and automated methods.

We first collected two lists of manually constructed labelings of news websites, one from [13] and the other from [40]. Grinberg et al. [13] provide a list of 1,253 news websites and 298 non-news websites that represented approximately 80% of all exposures to any form of content in their data. The list from [40] is a set of 7,845 news websites and is an aggregation of several hand-labeled datasets of local news websites available on the web. In addition to the use of these lists of domains, we also manually labeled the

[1] https://www.projectinfolit.org/

top 250 websites in our dataset (according to the number of unique URLs shared) as either "news" or "not news" domains. Using this combination of manually constructed lists from other sources and our own manually constructed list, we are able to classify 73% of all shares in our dataset as either "news" or "not news".

In order to better study the tail of the sharing distribution it was necessary to develop automated methods as well. We developed a deterministic[2] classifier by querying the Google News search index. We leveraged Google News because it represents a global independent standard for inclusion and exclusion of news, and it is a widely used aggregator of media content that combines human selection (outlets often have to apply for inclusion) and algorithmic curation.[3]

For each domain in our dataset that is shared more than once and was not already manually labeled (n=15,296), we performed a search on Google News with that domain name as the search term. We then recorded if any of the search results returned are a link to that domain.[4] If so, we classified the site as being a news domain, since it is indexed by Google News, otherwise, we assume the link is to a non-news website. We evaluated this approach by subsampling a random set of 275 websites from our data and manually classifying them as "news" or "not news". The classifier had an accuracy of 77.6%, acceptable for use in our labeling task. Using the Google News classifier, we were able to increase to 94% the number of shares in our dataset annotated as news or not news.

3.3 Identifying Latent Factors

Because our survey data was not constructed with an a priori expectation about factors indicative of breaking news sharing, we identify such factors using an exploratory method. To do so, we first identified within the survey a set of 28 questions that were potentially relevant to factors related to breaking news sharing. We then estimated an exploratory multiple item response theory model (MIRT) to identify latent factors (or factors) that drove patterns in responses, and how much each survey respondent was associated with each of those factors.

We describe here five sets of questions we used at a high level, the full set of questions that factor into our analyses are presented in Figure 1. The first set was a collection of 11 questions that attempted to establish why students shared breaking news specifically. Respondents were asked: "When you're deciding to share 'breaking news'–a special news event that is currently developing–on social media how do you evaluate the quality of the information that you share, if you do at all?" Respondents were then prompted with statements like: "Compare and fact check the news item using a different source," and asked on a five-point Likert scale from "Almost Always" to "Never" how often they used this behavior to evaluate information quality. Students could also answer: "I don't share breaking news at all," a point we address below.

The second set was a collection of eight questions that addressed at why students shared news in general. Students were asked, "Why do you share news items, if at all, on the social media sites that you use?", and then prompted to describe how strongly they agreed or disagreed with statements like: "Sharing news lets my friends/followers know about something I think they should know." Responses were on a five-point Likert scale from "Strongly Agree" to "Strongly Disagree", where students could also answer "I don't share or retweet news items at all."

The third, fourth, and fifth sets related to information veracity and fake news, underscoring the relationship between breaking news and misinformation noted above. The third set was a collection of six questions that attempted to determine students' perspectives on the factual nature of news. Students were prompted with statements like: "From my perspective, I do not trust the news, no matter what the source is." They were then asked to agree or disagree on the same five point Likert scale described above. The fourth set was a pair of questions that, as part of a broader question on what constitutes news, asked students to agree or disagree with the following two statements: "It's difficult to tell real news from fake news", and "The sheer amount of news on any given day is overwhelming." The final set was a single question that asked students to rate themselves on a five-point Likert scale ranging from "Very confident" to "Not confident at all" on the following question: "'Fake news' is a term we hear and see a lot these days. How confident do you feel with recognizing fake news?"

Once we identified the questions of interest, we used the `nFactors` package in R [28] to compare various heuristics for determining the number of factors to extract from the data. Using the Eigenvalues, Parallel Analysis, and Optimal Coordinates approaches, we settled on the extraction of seven latent factors from the responses to these questions. Because we expected there would be multiple factors to analyze, and because the data was represented via Likert scales, we then estimated a MIRT model using the `mirt` package in R [6]. This model provided us with a set of seven latent factors that best explain the variance in responses to the survey questions.

In the MIRT model each question "loads onto" (is associated with) zero or more factors or factors. Responses to those questions then determine the respondent's level for each of the factors. The set of latent factors is therefore essentially a smaller dimensional representation of the data; these dimensions describe a common set of factors that can be said to have driven responses to the survey data. The MIRT model we use is therefore a latent variable model that attempts to identify k (here, $k = 7$) latent continuous factors that can be used to maximally explain the variation in a set of ordinal responses. In comparison to similar approaches using matrix decomposition (e.g., with SVD) or Bayesian admixtures (e.g., latent Dirichlet allocation), the MIRT model is useful in that it can easily handle both missing data and ordinal data.

Because the model cannot be solved exactly, estimation was performed using a Monte-Carlo Expectation Maximization algorithm, as suggested by the package developers. The estimation procedure was run for 500 iterations, after which we extracted factors using an Oblimin rotation, ignoring loadings smaller than .25 on each dimension.

One final note regarding the questions "I don't share breaking news at all" and "I don't share or retweet news items at all". These accounted for approximately one third of our data for each question type. First, we required that these responses be consistent across entire question subsets (i.e. students answered either "I don't share

[2]In the sense that we do not train its parameters, although there is likely some stochasticity to the Google News search results
[3]See https://support.google.com/news/publisher-center/answer/7526416?hl=en.
[4]Note that while this approach worked in April of 2018, Google has since changed their site to use internal links, making this approach significantly more difficult.

breaking news at all" for each question or gave a Likert response for every question in the first set of questions). Second, because the MIRT model does not require non-missingness (it calculates likelihood estimates only over non-missing values), we treated cases where respondents gave these answers as missing in the MIRT model. Finally, we include. as a variable in the regressions described below, indicators for whether or not students provided these two responses.

3.4 Associating Latent Factors with Sharing Patterns on Twitter

We carry out three separate regression models to better understand patterns in sharing on Twitter and to contextualize those results with patterns for news sharing in general, and the sharing of all other non-news content. Values for these dependent variables were computed by aggregating the number of breaking news, news, and not-news shares for each individual to get a count for each user over all of their tweets.

The primary independent variables we study are the seven latent factors identified for each individual in the survey data. We also include control variables for gender and political party, as well as three different measures of activity. Political party is coded as a categorical variable with five levels: very conservative, conservative, moderate, liberal, and very liberal. We control for two measures of self-reported activity by creating indicators variable showing whether or not the student stated: "I don't share breaking news at all" or "I don't share or retweet news items at all" on the surveys, respectively. We also control for a behavioral measure of general activity online, including as a factor the logarithm of the number of statuses we collected for the individual (plus a constant of one to avoid undefined values). All independent variables are centered and scaled by one standard deviation.

Since the dependent variables are count-distributed, we use a negative binomial regression model. Further, given the exploratory nature of our work, rather than assuming the full model to be the single best model, we instead estimate the power-set of all potential models using the `glmulti` package in R and provide coefficient estimates averaged over these models. Doing so provides greater certainty on our estimates across a range of potential modeling decisions [4].

4 RESULTS

We first present results for our exploratory factor analysis generated by our MIRT model, which exposed seven latent factors that helped to explain responses to the survey questions detailed above. We then provide results for our regression analyses. These results motivated a final analysis that connects the latent factors we observe with partisan affiliation.

4.1 Analysis of Latent Factors

Figure 1 presents the seven latent factors that we identified from our MIRT model, and suggests several different ways in which responses to the survey questions are structured. Before continuing, we describe how to interpret these results by rough analogy to a model computational social scientists are likely to be more familiar with - latent Dirichlet allocation (LDA; [3]). Like LDA, our MIRT model decomposes the question responses into a set of latent factors ("topics" in LDA) onto which one or more questions (words, in LDA) are "loaded" (in LDA, that have a high probabilistic association with a topic). We can then use these latent factors, in combination with survey responses, to characterize individuals with a low-dimensional representation. This is analogous to LDA where we use topic distributions of documents as a representation of the content of that document.

In light of this explanation, Factor 1 represents students who find news overwhelming and who do not have confidence that they can differentiate real from fake news. Students who load high on Factor 1 are therefore considered apprehensive news sharers. Students who load high on Factor 2 can be said to have high trust in traditional media; meaning that they tend to trust news provided that it comes from particular sources, believe journalists have good intentions, and look for news from sources they trust. Factor 3 also represents trust, but in a different way. Students who load high on Factor 3 place less trust in specific media organizations (they do not find source to be a credibility indicator) but trust the news media writ large, especially with respect to particular journalists.

Students who loaded high on Factor 4 seemed to use both high and low quality indicators of source credibility. These individuals can be characterized as cautious in a different way than Factor 1; rather than being generally skeptical of their ability to identify real from fake news, and feeling overwhelmed, these individuals took clear steps to try to identify such differences. Students loading high on Factor 4 therefore actively worked to address potential issues with news content.

Students loading high on Factor 5 also spend time attempting to evaluate source credibility, but tend to do so with more social signals - evaluating hashtags, numbers of shares, and looking at the comments. Finally, students loading high on Factors 6 and 7 have different reasons for sharing. Students loading high on Factor 6 seem to dislike the idea of sharing in general, whereas students loading high on Factor 7 seem to be opposed only to their idea that their sharing online was an act intended to change the views of their peers or to give them a broader voice.

In sum, our MIRT model exposed latent factors identifying various ways in which students were cautious (or not) in evaluating news content before sharing, as well as different reasons they chose to share online. We now contextualize these factors by showing how they relate to real-world sharing behaviors of breaking news, news in general, and non-news content.

4.2 Connecting Latent Factors to Sharing Patterns

Figure 2 shows that even when controlling for gender, party, and various measures of self-reported and actual activity, the latent factors identified in our MIRT model help to explain sharing behavior on Twitter. The figure provides coefficient estimates with 95% confidence intervals of Incident Rate Ratios (IRR), i.e., the multiplicative change in the number of shares due to a one standard deviation increase in the variable. Results are shown for each of our three different dependent variables: breaking news (black), news sharing in general (orange), and non-news sharing (purple).

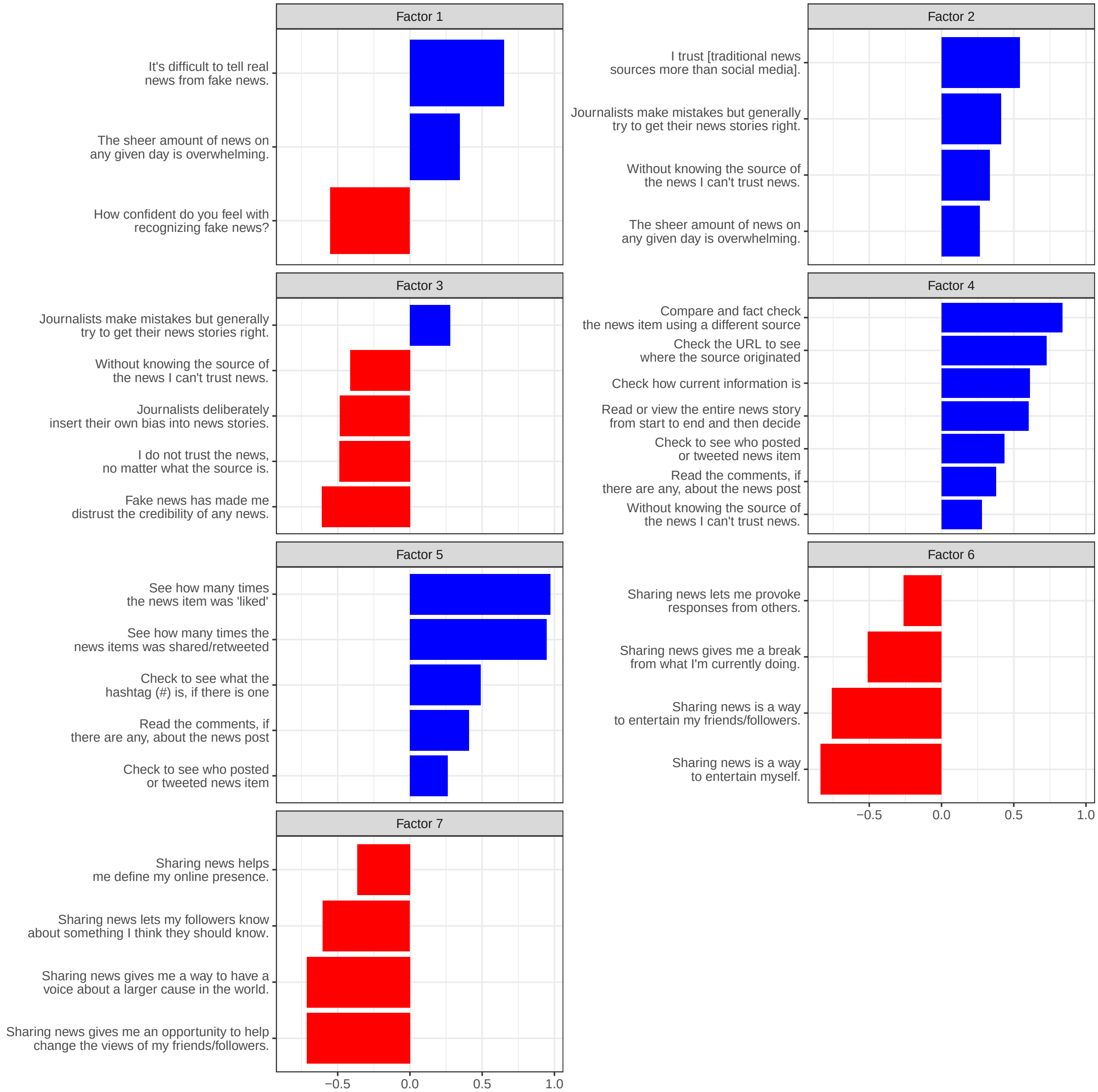

Figure 1: Results from the Multiple Item Response Theory Model. Each of the seven subplots shows a different latent factor identified by the model. For each factor, we show the questions that load most heavily onto it, as well as if that loading was positive (blue bar) or negative (red bar). For example, individuals who strongly agreed with the questions "It's difficult to tell real news from fake news" and "The sheer amount of news on ...", and who felt extremely unconfident recognizing fake news would have a strong positive association with Factor 1.

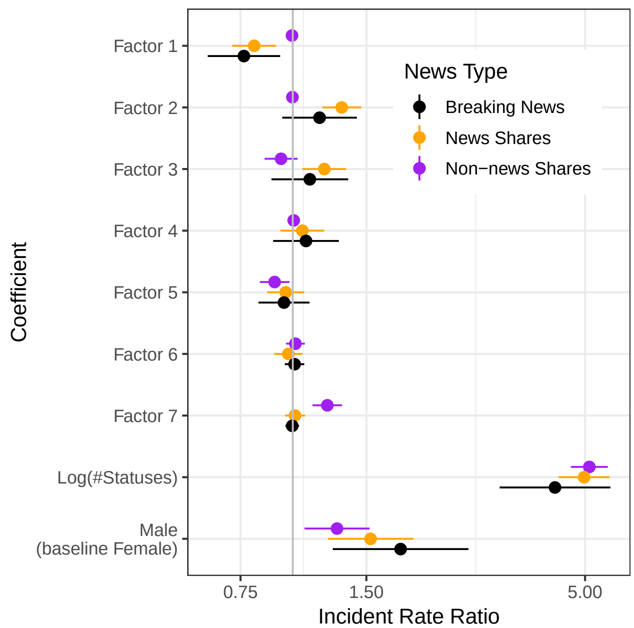

Figure 2: Results for our negative binomial regression models on the sharing of breaking news (black), all news (orange), and non-news (purple) shares. Results are given as 95% confidence intervals of Incident Rate Ratios (y-axis) for variables in the model (x-axis). The grey bar at 1 indicates no change from the variable. Not shown are coefficients for the two survey-based activity measures, or for the party variables, in order to ease visual clarify. Results for these coefficients are discussed in the text.

With respect to the latent factors associated with sharing on Twitter, Figure 2 shows that students who were unsure that they could differentiate between real and fake news, and who felt overwhelmed by the news, tended to share approximately 75% as much breaking news or news in general. In other words, students who were *more* certain of their ability to differentiate real and fake news, and who felt *less* overwhelmed, shared 25% *more* breaking news and news in general than their peers.

However, there was no association between breaking news sharing and Factor 4, which represented students who reported using a variety of different strategies to assess the veracity of news content. Consequently, it was students who believed they could differentiate real versus fake news, but not necessarily those who actually engaged in strategies to do so, who shared more breaking news. With respect to news sharing in general, students who placed high levels of trust in both the media writ large and in specific sources were likely to share more. Thus, general news sharing was associated with strong trust in traditional media and journalists and in oneself to identify such sources. Importantly, these behaviors were distinct from non-news sharing, which was significantly associated only with the use of low-quality credibility assessments of news and more individualistic reasons behind motivations for sharing.

With respect to control variables, men tend to share most across all three forms of sharing. And, as expected, individuals who are more active on Twitter tend to share more - a one standard deviation in the logged activity count was associated with almost five times more sharing of any kind. However, after controlling for this behavioral measure, self-reported measures of activity showed almost no association with sharing behavior. The single exception was that a small effect of stating: "I do not share or retweet news items at all", which was associated with a 40% drop in news sharing (95% IRR CI [.40, .97]). Finally, the categorical party affiliation variable is not a significant predictor of sharing behavior of any kind.

4.3 Connecting Latent Factors to Partisanship

Given established differences in the literature on sharing patterns across party lines, especially of misinformation [13, 14], we chose to further explore how party might be associated with sharing behavior through its relation to our latent factors. Figure 3 presents mean loadings, with confidence intervals, for individuals self-identifying with the five potential responses for party affiliations and their mean loadings on each factor.

Figure 3 shows that individuals who were more extreme in their party affiliations tended to load lower on Factor 1 - that is, they tend to be more confident in their ability to manage information overload and to differentiate real from fake news. While prior work has shown that individuals who are more avid news consumers also tend to be more politically active [7], other researchers find using behavioral data that individuals on the extreme right shared most fake news [13]. Given the lack of association between party and Factor 5 — the factor associated with use of fact-checking strategies— one hypothesis for future work that is consistent with our data is that people on the political extremes believe themselves to be knowledgeable news consumers, so much so that they do not necessarily engage in fact-checking behaviors that may prevent the spread of misinformed breaking news.

In addition, we find that Factors 2 and 3, which areassociated with trust in the media, are also associated with a left-leaning political affiliation. Consequently, an increase in traditional news media sharing is associated with an increase in media trust, which is in turn associated with a left-leaning political view. This finding fits with critiques of a left-leaning skew of traditional news media identified elsewhere [38].

In sum, party affiliation is not, controlling for other factors, significantly associated with sharing of breaking news, news in general, or non-news content. However, it is associated with latent factors, or traits, of individuals which do predict news sharing. These findings suggest the importance of continuing work to better understand the relationships between ideology and news sharing, particularly in the context of misinformation [16]. For instance, if such relationships exist at a cognitive level, or if the current political climate is more at fault for current political asymmetries.

5 CONCLUSION

As the media world grows more complex, explaining how people are exposed to news becomes more difficult [39]. Media content may be increasingly personalized and tailored to individuals by their own

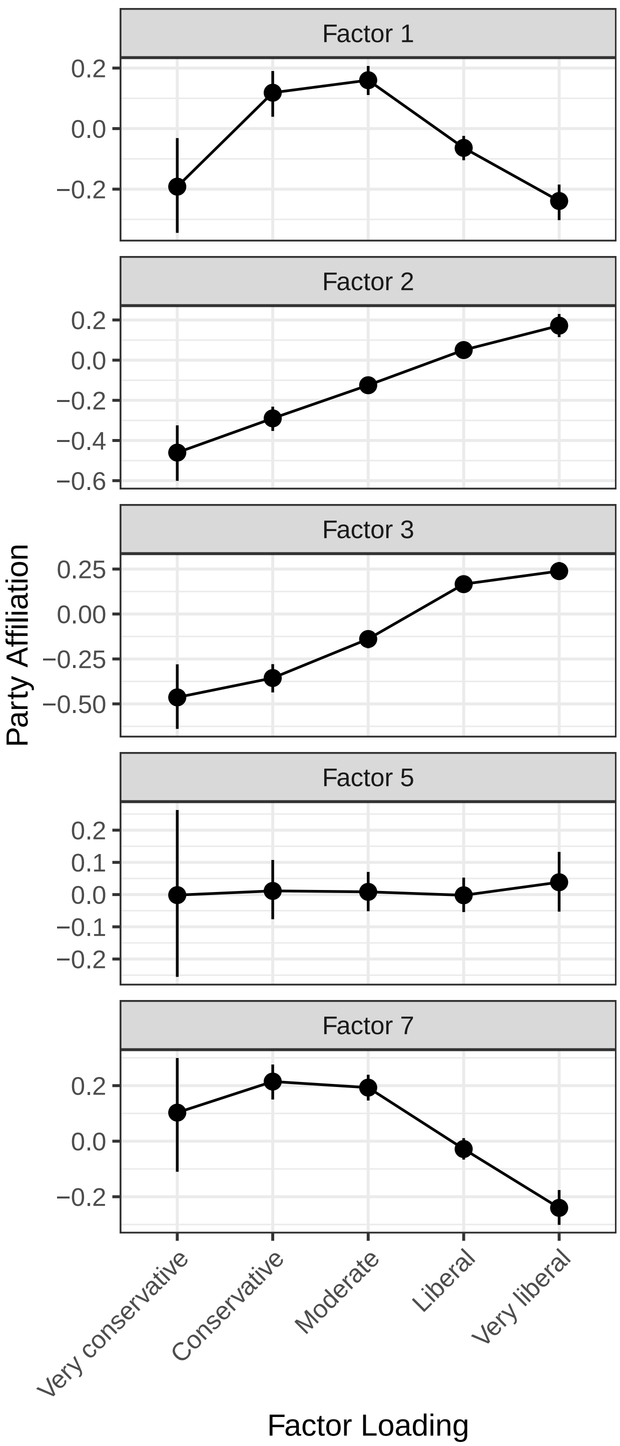

Figure 3: Average loading for each factor (y-axis) for individuals in the full survey dataset across party (x-axis). Each factor is a separate subplot, shown only are those factors that have a significant association to one of the sharing counts (breaking, news in general, or non-news). Confidence intervals shown for each party affiliation/latent factor combination are 95% confidence intervals.

choices and those of algorithms. Knowing how and why individuals play an active, mediating role in their online social environment is a vital step toward building better models for understanding the hybrid media ecosystem that characterizes 21st century societies [34].

The current paper makes a contribution to this area of inquiry by providing a rich, granular account of how and why members of the Millennial and Generation Z cohorts share breaking news content, and contextualizes the results with findings for news in general and non-news content. We do this by leveraging survey data and online observational data that are linked together for a large group of individual young persons.

Importantly, this study is based on a non-probability sample of young persons, and it is unclear how generalizable the findings may be. However, significant effort was made during the data collection phase to select a diverse group of colleges and universities across the United States (in terms of size, geography, and competitiveness of admissions) from which to draw a sample. Additionally, the survey data we rely on is self-reported, with all of the typical limitations of such data. To address these limitations, we use behavioral data to measure our dependent variable. Finally, we use some automation and heuristics to identify both breaking news and news in general, and an exploratory model of latent traits associated with news sharing, any of which, if modified, could potentially lead to different conclusions. However, the methods we use to identify news are relatively straightforward, and our statistical approach is well established in the survey literature.

Overall, we find that individuals' self-reported level of confidence in determining valid information from misinformation predicts increased sharing of breaking news content. However, these same individuals did not claim to engage in typical strategies for actually checking the veracity of breaking news content. Consequently, we find that individuals who believe themselves to be news "experts" share more breaking news, potentially irrespective of the aptitude of this self-identified expertise. Respondents who share more overall news seem to be more conscious of the source of content and generally trust news media more, but similarly did not necessarily engage in fact-checking strategies.

While these findings can clearly be cast in a negative light, there is also room for optimism. Specifically, there is reason to believe that young people do, in fact, have the news literacy they claim to [8], and it is simply expressed in ways that are not captured by the behaviors asked about in this survey. Whether or not these young, active sharers of breaking news are actually savvy news consumers is a matter of future work.

REFERENCES

[1] Karine Barzilai-Nahon. 2008. Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control. *Journal of the American Society for Information Science and Technology* 59, 9 (2008), 1493–1512.
[2] W. Lance Bennett and Alexandra Segerberg. 2012. The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics. *Information, Communication & Society* 15, 5 (2012), 739–768.
[3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022.
[4] Vincent Calcagno, Claire de Mazancourt, et al. 2010. glmulti: an R package for easy automated model selection with (generalized) linear models. *Journal of statistical software* 34, 12 (2010), 1–29.
[5] Manuel Castells. 2013. *Communication Power.* OUP Oxford.

[6] R Philip Chalmers et al. 2012. mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software* 48, 6 (2012), 1–29.

[7] Jihyang Choi, Jae Kook Lee, and Emily T. Metzgar. 2017. Investigating Effects of Social Media News Sharing on the Relationship between Network Heterogeneity and Political Participation. *Computers in Human Behavior* 75 (2017), 25–31.

[8] Stephanie Craft, Seth Ashley, and Adam Maksl. 2016. Elements of News Literacy: A Focus Group Study of How Teenagers Define News and Why They Consume It. *Electronic News* 10, 3 (2016), 143–160.

[9] Stine Eckert, Pradeep Sopory, Ashleigh Day, Lee Wilkins, Donyale Padgett, Julie Novak, Jane Noyes, Tomas Allen, Nyka Alexander, and Marsha Vanderford. 2018. Health-Related Disaster Communication and Social Media: Mixed-Method Systematic Review. *Health communication* 33, 12 (2018), 1389–1400.

[10] Deen Freelon and David Karpf. 2015. Of Big Birds and Bayonets: Hybrid Twitter Interactivity in the 2012 Presidential Debates. *Information, Communication & Society* 18, 4 (April 2015), 390–406.

[11] Jeffrey Gottfried and Elizabeth Grieco. 2018. Younger Americans Are Better than Older Americans at Telling Factual News Statements from Opinions.

[12] Jeffrey Gottfried and Elisa Shearer. 2017. Internet Closes in on TV as a Source of News in U.S. http://www.pewresearch.org/fact-tank/2017/09/07/americans-online-news-use-vs-tv-news-use/.

[13] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake News on Twitter during the 2016 U.S. Presidential Election. *Science* 363, 6425 (Jan. 2019), 374–378.

[14] Andrew Guess, Kevin Munger, Jonathan Nagler, and Joshua Tucker. 2018. How Accurate Are Survey Responses on Social Media and Politics? *Political Communication* 0, 0 (Nov. 2018), 1–18.

[15] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook. *Science Advances* 5, 1 (Jan. 2019), eaau4586.

[16] John T Jost, Sander van der Linden, Costas Panagopoulos, and Curtis D Hardin. 2018. Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current opinion in psychology* 23 (2018), 77–83.

[17] Anna Sophie Kümpel, Veronika Karnowski, and Till Keyling. 2015. News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Social Media+ Society* 1, 2 (2015), 2056305115610141.

[18] David Lazer. 2015. The Rise of the Social Algorithm. *Science* 348, 6239 (2015), 1090–1091.

[19] David MJ Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, and David Rothschild. 2018. The Science of Fake News. *Science* 359, 6380 (2018), 1094–1096.

[20] Chei Sian Lee and Long Ma. 2012. News Sharing in Social Media: The Effect of Gratifications and Prior Experience. *Computers in human behavior* 28, 2 (2012), 331–339.

[21] Adam Maksl, Seth Ashley, and Stephanie Craft. 2015. Measuring News Media Literacy. *Journal of Media Literacy Education* 6, 3 (March 2015), 29–45.

[22] Adam Maksl, Stephanie Craft, Seth Ashley, and Dean Miller. 2017. The Usefulness of a News Media Literacy Measure in Evaluating a News Literacy Curriculum. *Journalism & Mass Communication Educator* 72, 2 (June 2017), 228–241.

[23] Sharon Meraz and Zizi Papacharissi. 2013. Networked Gatekeeping and Networked Framing on #Egypt. *The International Journal of Press/Politics* 18, 2 (April 2013), 138–166.

[24] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. Tweeting Is Believing?: Understanding Microblog Credibility Perceptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM, 441–450.

[25] Johnny Nhan, Laura Huey, and Ryan Broll. 2017. Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings. *The British journal of criminology* 57, 2 (2017), 341–361.

[26] Thomas E. Patterson. 2000. *Doing Well and Doing Good: How Soft News and Critical Journalism Are Shrinking the News Audience and Weakening Democracy-and What News Outlets Can Do about It.* Joan Shorenstein Center on the Press, Politics and Public Policy, John F. Kennedy School of Government. Harvard University.

[27] Sasa Petrovic, Miles Osborne, Richard McCreadie, Craig Macdonald, Iadh Ounis, and Luke Shrimpton. 2013. Can Twitter Replace Newswire for Breaking News?. In *ICWSM*.

[28] Gilles Raiche, David Magis, and Maintainer Gilles Raiche. 2010. Package 'nFactors'. *Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test. Available at CRAN repository: http://cran. r-project. org/web/packages/n Factors/n Factors. pdf* (2010).

[29] Shafiza Mohd Shariff, Xiuzhen Zhang, and Mark Sanderson. 2017. On the Credibility Perception of News on Twitter: Readers, Topics and Features. *Computers in Human Behavior* 75 (2017), 785–796.

[30] Kate Starbird and Leysia Palen. 2011. "Voluntweeters": Self-Organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1071–1080.

[31] Alexandra Stevenson. 2018. Facebook Admits It Was Used to Incite Violence in Myanmar. *The New York Times* (Nov. 2018).

[32] Sue Yeon Syn and Sanghee Oh. 2015. Why Do Social Network Site Users Share Information on Facebook and Twitter? *Journal of Information Science* 41, 5 (2015), 553–569.

[33] Edson C. Tandoc Jr and Erika Johnson. 2016. Most Students Get Breaking News First from Twitter. *Newspaper Research Journal* 37, 2 (2016), 153–166.

[34] Kjerstin Thorson and Chris Wells. 2016. Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory* 26, 3 (Aug. 2016), 309–328.

[35] Jason Turcotte, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree. 2015. News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication* 20, 5 (2015), 520–535.

[36] Chris Wells and Kjerstin Thorson. 2017. Combining Big Data and Survey Techniques to Model Effects of Political Content Flows in Facebook. *Social Science Computer Review* 35, 1 (2017), 33–52.

[37] Stefanie Wiegand and Stuart E. Middleton. 2016. Veracity and Velocity of Social Media Content During Breaking News: Analysis of November 2015 Paris Shootings. In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 751–756.

[38] John Wihbey, Kenneth Joseph, and David Lazer. 2018. The Social Silos of Journalism? Twitter, News Media and Partisan Segregation. *New Media & Society* (Oct. 2018), 1461444818807133.

[39] John P Wihbey. 2019. *The Social Fact: News and Knowledge in a Networked World*. Mit Press.

[40] Leon Yin. 2018. Local News Dataset. https://doi.org/10.5281/zenodo.1345145