Predicting Audience Engagement Across Social Media Platforms in the News Domain

Kholoud Khalil Aldous[1(✉)], Jisun An[2], and Bernard J. Jansen[2]

[1] College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
kaldous@mail.hbku.edu.qa

[2] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
{jisun.an,jjansen}@acm.org

Abstract. We analyze cross-platform factors for posts on both single and multiple social media platforms for numerous news outlets to better predict audience engagement, precisely the number of likes and comments. We collect 676,779 social media posts from 53 news outlets during eight months on four social media platforms (Facebook, Instagram, Twitter, and YouTube), along with the associated comments (more than 31 million) and the number of likes (more than 840 million). We develop a framework for predicting the audience engagement based on both linguistic features of the post and social media platform factors. Among other findings, results show that content with high engagement on one platform does not guarantee high engagement on another platform, even when news outlets use similar cross-platform posts; however, for some content, cross-sharing posts on a platform will increase overall audience engagement on another platform. As one of the few multiple social media platform studies, the findings have implications for the news domain, as well as other fields that distribute online content via social media.

Keywords: Audience engagement · News outlets · Social media

1 Introduction

News outlets that produce digital content rely on social media platforms for dissemination [34]. Nearly all news outlets engage in online content dissemination [13,26], that almost all have accounts on multiple social media platforms [37], and 67% of news readers get their news, at least in part, from social media platforms [35]. Given the importance of social media for reaching audiences, news outlets are interested in measuring and evaluating performance via these channels. As such, there are various audience engagement metrics with many of these metrics being platform-specific, such as the number of likes, and comments, which we frame into four levels, from private to public expressiveness of user engagement [3]. Additionally, there has been prior work that investigates audience engagement on various social media platforms from the perspective of the news outlet [5,25], the audience of a news outlet [38], or both [3].

I. Weber et al. (Eds.): SocInfo 2019, LNCS 11864, pp. 173–187, 2019.
https://doi.org/10.1007/978-3-030-34971-4_12

Although impactful, much of this prior work focused on single news outlets (e.g., The New York Times, Al Jazeera), single topics (e.g., disasters, sporting events), or single platforms (e.g., Twitter, YouTube). There is a critical need for an analysis of multiple news outlets across multiple social media platforms unconstrained by topic. It is obvious that news outlets do not operate in a vacuum and that many times, they are competing with other news and other content creation outlets across multiple online platforms for audience numbers or are sharing an audience with other content organizations (i.e., a reader engaging with the same content from multiple news sources). Therefore, to truly investigate audience engagement effects, one should not limit the investigation to a single news outlet or a single social media platform; the investigation requires multiple news outlets posting content across multiple social media platforms. Although there has been some prior work looking into multiple news outlets [5,38], this work does not focus on audience engagement. There is also a general scarcity of cross-platform social media studies, especially in the news domain. Although a few prior works have focused on more than one platform [7,11,27], little has focused on the news industry. This lack of cross-platform analysis is a particularly notable shortcoming, as studies highlight that people generally have multiple social media accounts and get their news from multiple sources [38].

Gottfried [16] reports that 34% of users access multiple social media accounts to consume news, and some of these social media platforms attract millions of users (i.e., audience members of the news outlets) in a given day with the largest social media platforms used by news outlet being Facebook, Twitter, Instagram, and YouTube [37]. Understanding platforms differences and audience preferences across platforms are two main challenges for creating engaging content [2]. Given that news outlets are employing multiple social media platforms to distribute content [37] and the audience they are attempting to engage with are interacting with multiple social media platforms and multiple news outlets [16,33], it seems reasonable that there may be interaction effects as news outlets attempt to engage with their audience across these platforms. However, given the lack of prior work with both multiple organizations and platforms, there are several open questions. *Does audience engagement differ across platforms? Are there different audience behaviors for a news outlet on different platforms? Is news content transferable across platforms with similar engagement?* These are some of the questions that motivate our research.

To address a portion of these questions, we analyze social media posts from 53 worldwide news outlets across four major social media platforms during a continuous 8-month period. The contributions of this research are, first, an inclusive social media analysis of audience engagement for multiple major news outlets rather than an investigation of a single news source. Second, an analysis of these news outlets across multiple social media platforms, which allows for both single and cross-platform effects. Finally, the research examines potential audience engagement differences across platforms by analyzing news content posted on multiple platforms, shedding light on possible audience engagement strategies for these news outlets as they operate in a competitive online environment.

Our research differs from the existing work in that we perform a cross-platform study for four of the most active social media platforms in the news domain. Second, we predict audience engagement (high or low) for a social media post before it is published by employing linguistic features of the posts and cross-platform engagement feature. Lastly, for researchers and practitioners, this research offers insights regarding how to conduct multi-platform studies and the relationship between platforms' affordances and audience engagement.

2 Research Objectives

We aim to understand what are the similarities and differences in news outlets posting behavior and the effect of cross-sharing engagement factors for improving the prediction of user engagement. Our results can assist news outlets when using multiple social media platforms for distributing content. Focusing on the linguistic and cross-sharing aspects of the social media posts, we formulate the following research questions (RQ):

RQ1: Do news outlets tailor content when posting the same story on multiple platforms?
RQ2: Does cross-platform posting contribute to higher audience engagement?

To answer the first question, we extract the content that news outlets posted to more than one social media platform, called *cross-shared posts*. Then, we use similarity metrics to analyze the wording differences of posts across platforms. In addressing the second research question, we use *cross-shared posts*, to investigate audience engagement (likes and comments). We build a model that predicts the level of audience engagement given to a post for each platform. We use two different feature sets: (1) language features as Term Frequency-Inverse Document Frequency (TF-IDF) matrix, and (2) the level of audience engagement on another platform (called cross-platform engagement feature). By comparing the two sets of features for these multiple platform content, findings shed light on the effects of platforms impacting audience engagement.

3 Related Work

3.1 Audience Engagement in the News Domains

Audience engagement is a broad concept that encompasses a variety of phenomenon, including exposure, attention, interaction, and involvement [21,30]. At its base, engagement begins with exposure, but it constitutes additional psychological and behavioral experiences [12,30] that can be viewed from the perspective of the user, the medium, or both [32]. However, research on audience engagement typically focuses on producing, consuming, interacting with, or disseminating information, and collecting metrics to measure these engagement types [22]. This implies that the audience takes some action with the content as a fundamental component of engagement. It is these actions that news outlets are interested in measuring on social media platforms. The expected benefits for news outlets from a high level of audience engagement are superior outcomes, including popularity growth, cost reduction, and brand referrals [29]. Research on news consumption has relied on measuring user behaviors, such as page views and likes, as a composite of various metrics of both exposure and engagement behaviors [18]. One of the reasons that news outlets are interested in distributing content via social media platforms is that, in addition to the massive potential for exposure, online delivery of media via these platforms permits measurement at the individual story level. Additionally, audiences that satisfied with how they used a social media platform continually used the platform. Social media platforms also provide rather precise ways to measure whether and how a specific piece of content was consumed and, to some degree, a level of audience engagement; although, there may be news outlet differences on attracting audience [20].

3.2 Cross-Platform Analysis of Audience Engagement

While there are some studies using datasets from multiple platforms [3,6,7], there has been little large-scale cross-platform social media analysis in any domain. Studies have mainly focused on user behaviors and user interests on different platforms [23,24,39]. Additionally, there have been some studies in the news domain [10,15,17,36]; however, they are small-scale studies. Research has examined cross-posting activity on multiple platforms on a larger scale, but it did not look at audience engagement [14]. Prior work [11] has shown that it is possible to accurately model overall traffic by observing the first few minutes of social media reactions. In [27], researchers report that the volume and attitude of social media posts have a significant causality relationship to the amount of web searching and that they have interaction effects on social media traffic [28], although, the focus of these studies was not specific to the news domain.

4 Data Collection and Methodology

To investigate our research questions, we develop a list (a) of news outlets and (b) the social media platforms on which these news outlets posted content. We then collect the content that these news outlets posted to the social media platforms, along with audience engagement numbers for each post. We describe our data collection methodology here.

4.1 Selection of News Outlets and Social Media Platforms

We identify English news outlets with popular online presences using different ranking sources of news sources, including PewResearch [1] and Wallethub [31]. After examining the social media presence of different news outlets, we keep those outlets that are active across different social media platforms. This process results in 53 news outlets (see Appendix A).

For this cross-social media platform research, we select the four most popular social media platforms used most by news outlet based on eBizMBA rank [19], which are Facebook, Twitter, Instagram, and YouTube. We then identify the verified social media profiles for each news outlet on each of these platforms. For each news outlet, we select one social media account per platform that has worldwide English news and not specialized for a topic (e.g., sports).

4.2 Data Collection

In creating our dataset, we target the 53 news outlets' social media accounts on each of the four platforms for an eight months data collection period, January through August 2017, inclusive. We report the total number of collected posts with the total associated engagement metrics (likes and comments) in Table 1. In YouTube, two news outlets were not active, and two others have disabled the commenting feature.

Facebook (FB): For collecting Facebook posts from the news outlet' pages, We build a web crawler via the Facebook API. Each collected post is associated with audience engagement metrics (e.g., number of likes and comments).

Instagram (IG): Using the profile name of each news outlet, we implement a crawler that retrieves all Instagram posts, with associated engagement metrics, that are publicly available online in their profile. Then, for only retaining the eight months posts, we filter them by posting time.

Twitter (TW): To overcome the limitations of Twitter's maximum number of posts that can be collected, from an account, we use a web scraper to collect all publicly available posts. We collect each news outlet's tweets IDs using the eight months time filter. Using the IDs of the tweets, we collect the content of the tweets with associated audience engagement metrics through Twitter API.

YouTube (YT): The search function of YouTube Data API takes the news outlet channel name as input and return the publicly available YouTube posts information of that news outlet. Using this function, we collect all videos posts information including title, description, and engagement metrics.

4.3 Engagement Metrics

We use two engagement metrics: (a) likes ratio (LR) and (b) comments ratio (CR), as 'likes' and 'comments' are common and measurable across all the four

Table 1. Count of social media posts, comments, and likes by platform and totals.

	Posts	Comments	Likes
FB	27,117	984,266	70,557,281
IG	35,289	11,732,837	723,493,279
TW	571,270	14,426,570	13,604,785
YT	43,103	4,674,630	33,265,610
Total	676,779	31,818,303	840,920,955

social platforms. We also considered share ratio (e.g., shares on FB or retweets on TW), but we found it highly correlated with LR (Pearson correlation of 0.83), so we do not report it in this work. Instead, we focus on analyzing user engagement at the post-level, using likes and comments.

Likes Ratio (LR): The number of likes is a common engagement metric for a post on many platforms. Our dataset contains posts that received thousands of likes; however, plenty of posts also received zero likes. In order to deal with the huge difference between the number of likes for different posts, we apply a log function. Furthermore, posts tend to receive likes over a specific period; to accommodate this effect, we divide the number of likes by duration in days for normalization. The duration is the number of days from posting time to the day we collected the data. Normalizing by the maximum number of likes is required for the news media of which the content is posted, as each news media has its maximum value. We consider this value as an estimation for the audience volume.

A limitation for this method is that some posts that are very close to the collection period can get more credit compared to the posts of the first day of the collection time, as some posts do keep getting likes and comments for a more extended period than others. However, this way of normalization helps on standardizing the measure across platforms. Throughout the paper, we use likes ratio to represent the log-normalized values for the number of likes. The equation used for calculating LR for a post i is the following: $LR_i = \log_2(\frac{L_i+1}{T_i+M_i})$, where L_i is the number of likes of the post i, and T_i is the number of days from the posting day till the collection day, inclusive. M_i is the maximum number of likes for the news outlet that posted the content.

Comments Ratio (CR): Similar to the likes ratio, we use the log-normalization approach to calculate the comments ratio using the posts' number of comments. The equation used for calculating CR for a post i is: $CR_i = \log_2(\frac{C_i+1}{T_i+M_i})$, where C_i is the number of comments and T_i is the number of days from the posting day till the collection day, inclusive. M_i is the maximum number of comments for the news outlet that posted the content.

4.4 Extracting Cross-Shared Post

We extract cross-shared posts, which is the content that news outlets have posted on more than one social media platform. To determine cross-shared posts, we pair posts from each platform with posts on the other three platforms for each news outlet within three days. For all potential post pairs, we adopt a content-based matching method used in prior work for finding the same user across social media platforms with a 94.5% accuracy [4] and an approach shown sufficient for news article [9]. In our case, we aim to find similar news posts across social media platforms. URLs matching is another possible way for finding cross-shared posts; however, URLs are not commonly used in Instagram and YouTube posts. Also, when expanding Twitter URLs, we found many invalid links, or they relate back to the homepage of the news outlet website and not to the specific news article.

Table 2. Number of cross-shared posts for platform pairs,with results of manual validation of 100 postings

Platform	#Pairs	Matched
FB-TW	5600	94%
FB-IG	3174	37%
FB-YT	5327	53%
IG-TW	10527	83%
IG-YT	8942	37%
TW-YT	2449	77%

To this end, we represent each post in a TF-IDF vector and then examine whether the two posts are similar or not by computing cosine similarity. We first apply tokenization and then construct a TF-IDF matrix for each posts pair. Once having the TF-IDF matrix, we compute the cosine similarity between two posts. We use a cosine similarity threshold of 0.3, eliminating all post pairs less than the threshold[1]. The result of this process is a set of post pairs for the six different platform pairs that are (a) Facebook-Twitter (FB-TW), (b) Facebook-Instagram (FB-IG), (c) Facebook-YouTube (FB-YT), (d) Instagram-YouTube (IG-YT), (e) Instagram-Twitter (IG-TW), (f) Twitter-YouTube (TW-YT). The number of pairs (#pairs) for each platform pair is shown in Table 2.

In order to validate the cross-shared posts for each platform pair, we manually labeled a random sample of 100 post pairs from each of the platform pairs. One of the authors performed the labeling–a pair of posts is labeled one (1) if they refer to the same story (cross-shared) and zero (0) otherwise. The findings from the manual validation match (matched) are shown in Table 2. We observe that the three pairs, including Twitter, FB-TW, IG-TW, and TW-YT, have high matching rates while the other three pairs have low matching rates. This result is in line with findings by [24], where Twitter was found to be the dominant destination for 54% of cross-sharing activities. One reason is that the cross-sharing support and usability within Instagram and YouTube use Twitter as a destination. Hence, the functional aspects of social platforms strongly affect the cross-sharing behavior. Based on the results, we use FB-TW, IG-TW, and TW-YT for our cross-shared posts analysis.

4.5 Building Models for Predicting Audience Engagement

We build a model that predicts audience engagement level of posts for each platform. We set our prediction task as a binary classification–the model predicts whether a given post will have high or low audience engagement on a platform, building separate models for each platform. We take the top and bottom 33% of

[1] The threshold was chosen based on our manual inspection of 600 randomly selected pairs.

the data and consider the top 33% as posts with high engagement and the bottom 33% as posts with low engagement. We use this method to distinguish highly engaging content from low engaging content and to make it clearer for the model for understanding the features differences. Another reason is to overcome the class imbalance problem, as having the top and bottom percentage equal (33%), the number of posts within each class is almost equal, and thus, the random baseline of our classifier is 0.5. Since each news media source has a different distribution for their LR and CR, we added news outlet as a categorical input feature to the model. Hence, we use different 0.33 and 0.66 percentile values across individual news media depending on their posts' engagement metrics. For example, on Twitter, the MSNBC news comments ratio top 33% is equal to 0.21 (0.66 percentile), and the bottom 33% value is -1.16 (0.33 percentile), so all posts with a comments ratio greater than 0.21 are labeled 1 (2,436 posts) and posts with comments ratio less than -1.16 are labeled 0 (2,348 posts). Posts in between are not considered in this analysis, and we reserve them for future research.

We use three different feature sets: (1) language features as a proxy of the topic of the content; (2) the level of audience engagement on another platform; and (3) news outlet as a categorical feature. To construct the language features for each platform, we take the text of all posts, and we remove punctuation and stop words, then we apply tokenization and stemming. Once we have cleaned the posts, we construct a TF-IDF matrix with setting three parameters: maximum IDF of 0.8, minimum IDF of 0.001, and the maximum number of features (i.e., words) of 2,000. The third feature is only available for cross-shared posts. Thus, we use this feature for examining audience engagement interaction among cross-shared posts. We use the news outlet categorical feature for building all models to address the effect of the actual news outlets.

To measure the prediction results, we use F1-score, Precision, Recall, and Area Under the Curve (AUC) using 10-fold cross-validation. We only report the F1 score, as results for all measures were positively correlated. We test several classification algorithms, including AdaBoost, Decision Tree, Logistic Regression, and Random Forest.

5 Results and Discussion

We aim to unveil whether knowing the level of audience engagement for a post on one platform can predict the audience engagement for the post on another platform. For this, we will use cross-shared posts (i.e., those posts a news outlet shares on more than one platform). We first explore the similarities between cross-shared posts.

5.1 Similarities of Cross-Shared Posts

For addressing the first research question (RQ1), we average three text similarity measures for the cross-shared posts, which are: Jaro distance, Levenshtein

Table 3. Average text similarity between cross-shared posts using three measures Jaro distance, Levenshtein distance, and cosine similarity. The last two columns show the average number of words per post for paired platforms (P1 and P2).

P1-P2	Jaro	Levenshtein	Cosine	Words (P1)	Words (P2)
TW-FB	0.75	81.00	0.52	116.00	135.66
TW-IG	0.71	227.69	0.51	108.39	284.60
TW-YT	0.59	601.23	0.38	119.72	681.20

distance, and cosine similarity of two posts. Jaro distance value range from 0 to 1, where the higher the Jaro distance for two posts is, the more similar the posts are. Levenshtein distance is the minimum number of single-character edits that are necessary to modify one post to obtain another post, hence the lower is the distance, the higher is the similarity between the two posts. Cosine similarity measures how similar the two posts are likely to be in terms of their subject matter, where a value of 1 means exact match and 0 means dissimilar. Table 3 shows the average similarities with the average number of words per post for both paired platforms posts. News outlets posting to TW and FB are the most similar with 0.75 similarity using Jaro distance. Then, TW and IG postings are less similar with Jaro distance of 0.71, which is related more to the extended size of IG posts, which is on average 284 word per post. The cross-shared posts between TW-YT are less similar with 0.59 Jaro and 601 Levenshtein distance.

Generally, news outlets tend to make posts different from each other when cross-posted on multiple platforms. It appears that some news outlets are making an effort to tailor their posts to each platform, even if the original article is the same. News outlets do tailor their posts to individual social media platforms, with varying degrees of similarity between platforms, which addresses the first research question.

5.2 Predicting Audience Engagement for All Posts

As explained in the methods section, we first predict the level of audience engagement for a given post for each platform using linguistic features. Table 4 presents the F1-scores of 10-fold cross-validation using AdaBoost, Decision Tree, Logistic Regression, and Random Forest on each platform. The best performing algorithm in this experiment is logistic regression for both LR and CR. The best F1 score is 0.69, which is the YouTube prediction model for CR using linguistic features (L). We can observe from this experiment that audience engagement can be predicted with F1 scores ranging from 0.62 to 69 across all platforms using logistic regression.

5.3 Predicting Audience Engagement for Cross-Shared Posts

Cross-shared posts provide a unique opportunity to understand how news outlet are leveraging multiple social media platforms in their audience engagement

Table 4. F1-scores of the four algorithms predicting engagement for a post on each platform using linguistic features (L).

	FB		IG		TW		YT	
	LR	CR	LR	CR	LR	CR	LR	CR
AdaBoost	0.45	0.54	0.66	0.49	0.52	0.59	0.56	0.61
Decision tree	0.59	0.62	0.63	0.60	0.62	0.65	0.61	0.61
Logistic regression	**0.62**	**0.64**	**0.68**	**0.66**	**0.65**	**0.68**	**0.67**	**0.69**
Random forest	0.60	**0.64**	0.63	0.61	**0.65**	0.67	0.61	0.63

strategies. We exploit the level of audience engagement on one platform in order to predict audience engagement on another platform. As mentioned in the methods section, we examined how many posts were cross-shared on two platforms (Table 2). We use the three pairs (FB-TW, IG-TW, and YT-TW), as they have the highest percentage of cross-shared posts (>75%).

Table 5 reports the Pearson's correlation coefficients for testing whether the level of audience engagement (either LR or CR) in two paired platforms are similar or not for those cross-shared posts. Facebook and Instagram have a positive correlation with Twitter for both LR and CR, indicating that a post having high LR/CR on Twitter is more likely to have high LR/CR on Facebook or Instagram, and vice versa. YouTube LR (CR) has a positive (negative) correlation with Twitter LR, but there is no significant relation between Twitter CR and YT. Overall, the results indicate that a post can be popular on different platforms, which could be the result of different factors, including posts' linguistic features or platforms' audience differences.

We build an audience engagement prediction model with the cross-shared posts for each of the platform pairs. A separate model for each platform, and thus, for example, the FB-TW pair results in two prediction models: (1) predicting audience engagement on FB using TW information and (2) predicting audience engagement on TW using FB information. Across all models, we use four algorithms (AdaBoost, Decision Tree, Logistic Regression, and Random Forest) and employ 10-fold cross-validation. Table 6 shows the results of predicting audience engagement on TW based on FB, IG, and YT information (LR or CR). We compare the model using linguistic features (L) to a model with an addi-

Table 5. The Pearson correlation between engagement metrics of cross-shared content based on cosine similarity.

		FB		IG		YT	
		LR	CR	LR	CR	LR	CR
TW	LR	0.27**	0.28**	0.52**	0.25**	0.36**	−0.23**
	CR	0.24**	0.19**	0.05**	0.18**	0.05*	0.01

Significant level codes: * $p < 0.05$, ** $p < 0.001$

Table 6. F1-scores of the four algorithms predicting engagement based on cross-shared posts on Twitter using other platforms information. Using linguistic features (L) and cross-platform information (C) with linguistic features (L+C).

			FB		IG		YT	
			L	L+C	L	L+C	L	L+C
TW	AdaBoost	LR	0.55	0.73	0.65	0.72	0.59	0.68
		CR	0.58	0.72	0.53	0.76	0.66	0.73
	Decision tree	LR	0.65	**0.74**	0.69	**0.75**	**0.71**	**0.71**
		CR	**0.68**	**0.74**	**0.70**	**0.77**	**0.73**	**0.77**
	Logistic regression	LR	**0.66**	0.70	**0.70**	**0.75**	0.69	0.69
		CR	**0.67**	0.69	**0.70**	0.74	0.72	0.72
	Random forest	LR	0.53	0.66	0.55	0.63	0.52	0.58
		CR	0.48	0.67	0.46	0.75	0.51	0.54

Table 7. F1-scores of the four algorithms for predicting engagement on the three platforms using Twitter information of cross-shared posts.

			FB		IG		YT	
			L	L+C	L	L+C	L	L+C
TW	AdaBoost	LR	0.48	0.73	0.70	0.72	0.59	0.64
		CR	0.63	0.73	0.65	0.75	0.65	0.67
	Decision tree	LR	**0.70**	**0.75**	**0.73**	**0.77**	**0.70**	**0.71**
		CR	**0.72**	**0.75**	**0.75**	**0.79**	**0.73**	0.71
	Logistic regression	LR	0.66	0.69	0.70	0.74	0.67	0.68
		CR	0.70	**0.75**	0.70	0.78	0.72	**0.72**
	Random forest	LR	0.52	0.66	0.67	0.69	0.54	0.58
		CR	0.54	0.67	0.62	0.73	0.53	0.56

tional feature of cross-platform engagement (C), i.e., the LR or CR on another platform. We denote this full model as L+C. On all platforms, using the cross-platform engagement feature significantly improves the F1-score by 7–23% for AdaBoost and 3–29% for Random Forest. Using logistic regression, LR and CR improve across FB and IG but not YT. Decision Tree shows the best F1-scores with an improvement of 6–9% when using L+C for both FB and IG, but shows slight improvement for YT CR and no effect using YT LR. Generally, there is certainly a cross-platform effect occurring, where the effect is more pronounced with FB and IG.

Table 7 shows the results of audience engagement prediction on the three platforms: FB, IG, and YT based on information from TW. FB achieves 25% and 10% improvement using AdaBoost for LR and CR, respectively. IG, taking similar advantage of TW information, has a 4% improvement using Decision Tree for LR and CR. Twitter information has slightly less effect on the prediction results of YT, with LR (5%) and CR (1%) using AdaBoost. However, using TW engagement feature to predict YT engagement with Decision Tree shows a slight improvement on LR and negative effect on CR.

Overall, for answering RQ2, we observe that using the cross-platform information can improve audience engagement prediction by 1–25%. The cross-shared effect is less between TW and YT in relative to TW and the other platforms. One reason could be that audience base or their preferences for the content of the two platforms are too different, or the technology affordance differences of the platforms do not lend themselves to shared content.

5.4 Implications for News Outlets

Our research findings using cross-shared posts provide new insights for using multiple social media platforms in the news domain, with possible implications for other domains and online user measurement [8]. First, content with high engagement on one platform does not guarantee success on another platform, even when news outlets use similar posts across platforms. Secondly, there is significant, notable engagement improvement for content that is shared cross-platform among Facebook, YouTube, and Instagram interacting with Twitter, indicating that Twitter may be a bridge platform to audience segments on these other platforms. However, there seems to be a reciprocal effect with cross-shared posts, also boosting engagement on Twitter. Finding cross-shared posts between platforms (e.g., FB-IG) other than Twitter is needed, which can be done through a crowdsourcing labeling task. The trend might differ by domains, and conducting similar experiments using datasets of other domains is needed. For instance, this "bridging" can potentially help in the marketing domain to define which platform best fits a marketing campaign and whether cross-sharing the campaign content onto other platforms generates more audience engagement.

6 Conclusions and Future Research

Prior understanding of audience engagement factors across social media platforms for a given domain was based on a patchwork of different studies, done across individual platforms, and focusing on one to a small number of news outlets, often with small datasets. In our research, we use a large number of news outlets with a large number of posts and analyze audience engagement across multiple platforms in order to study the cross-platform engagement effect. From our findings, we presented both theoretical and practical implications that further academic research and provides actionable advice for content producers. As such, the research presented here expands the prior work on engagement by focusing on multiple news outlets within a single domain and content from these news outlets in a cross-platform perspective.

In term of limitations, we highlight two confounding factors that could influence study findings, which are promoted or boosted posts. For promoted posts, since promotion works only on one platform, one could imagine that social media managers may boost different posts on different platforms according to different strategies which could throw off the observed values of engagement. Although we attempted to control the audience differences between outlets, many other user factors can influence the results. Different user features (e.g., demographics) and the overlapping percentage between platforms might affect the results. Also, the quality of matching cross-shared posts needs to be improved through manual labeling of all posts and not only a sample of 100, as done in this study. Finally, there is a need to validate whether or not CR and LR are good metrics compared to other metrics, such as impressions or reach.

A The list of news outlets

The list of the 53 news outlets is shown in Table 8:

Table 8. List of 53 online news organizations

ABC News	Los Angeles Times	The New York Times
AlJazeera	Mail Online	The Verge
BBC News	Mic	The Wall Street Journal
Bleacher Report	MSNBC	The Washington Post
Bloomberg	National Public Radio	The Week
Boston.com	NBC News	The Blaze
Breitbart News Network	NDTV	TIME
Business Insider	New York Post	U.S. News & World Report
BuzzFeed	Newsweek	Upworthy
CBS News	NY Daily News	USA Today
Chicago Tribune	Qatar Tribune	VICE
CNBC	Salon	Vox
CNN	Slate	Xinhua News Agency
CNN digital Network	The Associated Press	
Examiner.com	The Atlantic Magazine	
Financial Times	The Boston Globe	
Forbes	The Daily Beast	
Fortune	The Economist	
Fox News	The Guardian	
Huffington Post	The Hill	

References

1. The Top 25 (2011). https://www.journalism.org/2011/05/09/top-25/. Accessed 15 Apr 2019
2. Aldous, K.K., An, J., Jansen, B.J.: The challenges of creating engaging content : results from a focus group study of a popular news media organization. In: ACM CHI (2019)

3. Aldous, K.K., An, J., Jansen, B.J.: View, like, comment, post: analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In: ICWSM (2019)
4. Alvarez, J.J., Mendoza, F.A., Labrador, M.: An accurate way to cross reference users across social networks. In: SoutheastCon. IEEE (2017)
5. An, J., Cha, M., Gummadi, P.K., Crowcroft, J.: Media landscape in Twitter: a world of new conventions and political diversity. In: ICWSM (2011)
6. An, J., Quercia, D., Crowcroft, J.: Partisan sharing: Facebook evidence and societal consequences. In: COSN. ACM (2014)
7. An, J., Weber, I.: Diversity in online advertising: a case study of 69 brands on social media. In: Staab, S., Koltsova, O., Ignatov, D.I. (eds.) SocInfo 2018. LNCS, vol. 11185, pp. 38–53. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01129-1_3
8. Balbi, S., Misuraca, M., Scepi, G.: Combining different evaluation systems on social media for measuring user satisfaction. Inf. Process. Manag. **54**(4), 674–685 (2018)
9. Barrón-Cedeo, A., Jaradat, I., Martino, G.D.S., Nakov, P.: Proppy: organizing the news based on their propagandistic content. Inf. Process. Manag. **56**(5), 1849–1864 (2019)
10. Brems, C., Temmerman, M., Graham, T., Broersma, M.: Personal branding on Twitter. Digit. J. **5**(4), 443–459 (2017)
11. Castillo, C., El-Haddad, M., Pfeffer, J., Stempeck, M.: Characterizing the life cycle of online news stories using social media reactions. In: CSCW. ACM (2014)
12. Constantinides, M., Dowell, J.: A framework for interaction-driven user modeling of mobile news reading behaviour. In: UMAP. ACM (2018)
13. Dragoni, M., Federici, M., Rexha, A.: An unsupervised aspect extraction strategy for monitoring real-time reviews stream. Inf. Process. Manag. **56**(3), 1103–1118 (2019)
14. Farahbakhsh, R., Cuevas, A., Crespi, N.: Characterization of cross-posting activity for professional users across Facebook, Twitter and Google+. Soc. Netw. Anal. Min. **6**(1), 33 (2016)
15. Fletcher, R., Nielsen, R.K.: Are people incidentally exposed to news on social media? a comparative analysis. New Media Soc. **20**(7), 2450–2468 (2018)
16. Gottfried, J., Sheare, E.: News use across social media platforms 2016. State of the News Media, pp. 1–20 (2016)
17. Hladk, R., Štětka, V.: The powers that tweet. J. Stud. **18**(2), 154–174 (2017)
18. Hussein, R., Hassan, S.: Customer engagement on social media: how to enhance continuation of use. Online Inf. Rev. **41**(7), 1006–1028 (2017)
19. Kallas, P.: Top 15 most popular social networking sites, July 2017 (2017). http://www.ebizmba.com/articles/social-networking-websites. Accessed 19 Aug 2018
20. Kim, S.M., et al.: Twitter content eliciting user engagement: a case study on Australian organisations. In: WWW. ACM (2017)
21. Ksiazek, T.B., Peer, L., Lessard, K.: User engagement with online news: conceptualizing interactivity and exploring the relationship between online news videos and user comments. New Media Soc. **18**(3), 502–520 (2016)
22. Lalmas, M., O'Brien, H., Yom-Tov, E.: Measuring user engagement. Synth. Lect. Inf. Concepts, Retrieval Serv. **6**(4), 1–132 (2014)
23. Lee, R.K.W., Hoang, T.A., Lim, E.P.: On analyzing user topic-specific platform preferences across multiple social media sites. In: WWW (2017)
24. Lim, B.H., Lu, D., Chen, T., Kan, M.Y.: # mytweet via instagram: exploring user behaviour across multiple social networks. In: ASONAM. ACM (2015)

25. Mele, I., Bahrainian, S.A., Crestani, F.: Event mining and timeliness analysis from heterogeneous news streams. Inf. Process. Manag. **56**(3), 969–993 (2019)
26. Mitchell, A., Holcomb, J., Weisel, R.: Data and trends about key sectors in the U.S. news media industry. State of the News Media, pp. 1–118 (2016)
27. Mukherjee, P., Jansen, B.J.: Conversing and searching: the causal relationship between social media and web search. Internet Res. **27**(5), 1209–1226 (2017)
28. Mukherjee, P., Jansen, B.J.: Information sharing by viewers via second screens for in-real-life events. Trans. Web **11**(1), 1:1–1:24 (2017)
29. Muñoz Expósito, M., Oviedo-García, M., Castellanos-Verdugo, M.: How to measure engagement in Twitter: advancing a metric. Internet Res. **27**, 1122–1148 (2017)
30. Napoli, P.M.: Audience Evolution: New Technologies and the Transformation of Media Audiences. Columbia University Press, New York (2011)
31. Nicolae, M.: 2016's best news sites (2016). https://wallethub.com/blog/best-news-sites/21699/. Accessed 15 Apr 2019
32. O'Brien, H.L., Toms, E.G.: What is user engagement? a conceptual framework for defining user engagement with technology. Am. Soc. Inf. Sci. Technol. **59**(6), 938–955 (2008)
33. Picone, I., Courtois, C., Paulussen, S.: When news is everywhere: understanding participation, cross-mediality and mobility in journalism from a radical user perspective. J. Pract. **9**(1), 35–49 (2015)
34. Schlagwein, D., Hu, M.: How and why organisations use social media: five use types and their relation to absorptive capacity. J. Inf. Technol. **32**(2), 194–209 (2017)
35. Shearer, E., Gottfried, J.: News use across social media platforms 2017. Pew Research Center (2017)
36. Smith, A.N., Fischer, E., Yongjian, C.: How does brand-related user-generated content differ across Youtube, Facebook, and Twitter? J. Interact. Mark. **26**(2), 102–113 (2012)
37. Stocking, G.: Digital news fact sheet. State of the News Media, pp. 1–2 (2017)
38. Swart, J., Peters, C., Broersma, M.: Navigating cross-media news use: media repertoires and the value of news in everyday life. J. Stud. **18**(11), 1343–1362 (2017)
39. Zhong, C., Chan, H.w., Karamshuk, D., Lee, D., Sastry, N.: Wearing many (social) hats: how different are your different social network personae? In: ICWSM (2017)