Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)

View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations

Kholoud Khalil Aldous,[1] Jisun An,[2] Bernard J. Jansen[2]
[1]Hamad Bin Khalifa University [2]Qatar Computing Research Institute, Hamad Bin Khalifa University
kaldous@mail.hbku.edu.qa, jisun.an@acm.org, jjansen@acm.org

Abstract

We evaluate the effects of the topics of social media posts on audiences across five social media platforms (i.e., Facebook, Instagram, Twitter, YouTube, and Reddit) at four levels of user engagement. We collected 3,163,373 social posts from 53 news organizations across five platforms during an 8-month period. We analyzed the differences in news organization platform strategies by focusing on topic variations by organization and the corresponding effect on user engagement at four levels. Findings show that topic distribution varies by platform, although there are some topics that are popular across most platforms. User engagement levels vary both by topics and platforms. Finally, we show that one can predict if an article will be publicly shared to another platform by individuals with precision of approximately 80%. This research has implications for news organizations desiring to increase and to prioritize types of user engagement.

Introduction

News organizations rely on the major social media platforms to distribute content to their audiences. In addition to distributing content, social media platforms provide users with unprecedented means to express their reactions and interests on a wide range of issues (Thonet et al. 2017). More than half of news readers (67%) use social media to get at least part of their news, and one in four US adults get news from two or more social media platforms (Shearer and Gottfried 2017). Thus, social media is a major distribution venue for news organizations. Social media platforms are also an increasing source of revenue in which news organizations compete for a share of more than sixty billion spent yearly on digital advertising (Mitchell, Holcomb, and Weise 2017). Therefore, social media platforms are integral to both disseminating news content and generating revenue for news organizations (Mitchell, Holcomb, and Weise 2017). The challenges of creating engaging content include understanding audience preferences across platforms and between platforms (Aldous, An, and Jansen 2019). Consequently, understanding how users engage with content via social media platforms is critical for both the news organizations and the audiences that they serve, along with providing insights into the role of social media technology in this linkage.

Organizations that use social media platforms to disseminate content measure user engagement via various web analytics metrics. User engagement is a key performance indicator for digital organizations; user growth (e.g., followers, subscribers, etc.) and interactions (e.g., likes, comments, shares, etc.) are common measures of organizational success across social media platforms and indications of user satisfaction (Balbi, Misuraca, and Scepi 2018). Typically, the higher the volume of each of these metrics, the better.

However, not all user engagement metrics have the same impact, and different metrics indicate different types of user engagement and expression (Noguti 2016). Therefore, a standardized approach to categorizing them would benefit both organizations and social media platforms. One approach is to classify user engagement metrics into levels that represent similar actions, expressions, or impacts. The network effect is a key advantage of publishing content on social media platforms (Lin and Lu 2011). From the perspective of the content providers, the more public the expression of interest with the content by the user, the more potentially impactful that engagement is due to the network effect of social media platforms. Conceptually, this perspective is foundationally grounded in the concept of electronic word-of-mouth (eWOM), where users actively promote a product or service (Jansen et al. 2009). eWOM is extremely beneficial for news content providers as it is a measure of the content dissemination from the organizations' existing user base to a potentially greater audience at no cost to the organization. Therefore, the degree of public expression, from private to public, about content may be an effective way to organize user engagement metrics. Based on this conceptual background and building from (Noguti 2016), we derive a user engagement framework organized by the degree of user public expressiveness concerning a piece of content.

Table 1 presents four levels of user engagement with their definitions and example metrics. Each level indicates a given measure of engagement, with higher levels indicating a greater degree of public expressiveness by users. Lower levels indicate less public and more private expressions of engagement. Level-1 indicates a mostly private level of involvement by people, such as viewing content posted by others (e.g., viewing a video). Level-2 engagement involves people liking posted content, which presents their own preferences publicly (e.g., liking a video). Commenting on content is considered Level-3 of engagement, as a higher act of public expression than simply liking. The user engagement action of sharing a post on social media platforms we also consider Level-3, as it makes the content visible to the user's network on the platform. One would generally consider commenting or sharing as a more public act than liking, as they act to disseminate further the content on the platform, typically in a manner that others can easily observe.

Level	Definition	Example Metric
Level-1	Private engagement by viewing social media posts or videos	# views
Level-2	Exposing user preferences by liking social media posts	# likes
Level-3	Expressing opinion or feelings by commenting, sharing in private messages, or sharing on the same platform	# comments, # shares
Level-4	Spreading content through public sharing into other public networks or platforms	# external postings

Table 1: Engagement levels ordered by degree of public expressiveness, from more private to more public

Level-4 is the most public level of engagement and occurs when users include content from one social media network in their posts on other social media platforms. These re-postings onto different networks amplify the reach of the original content by disseminating it to a potentially wider user base. It enlarges the audience reach to a different network that differentiates Level-4 sharing from Level-3 sharing, which occurs on the same network as the original content posting. Although Level-4 engagement can occur on most social media services, Reddit is a good platform to study Level-4 public sharing because when individuals post to Reddit's different communities (e.g., r/worldnews, r/politics, r/todayilearned) the content's reach is enlarged to the new community subscribers (e.g., r/worldnews has 20.3 million subscribers). As such, we use Reddit in the research presented here. We highlight what we consider different engagement levels in each platform due to the platform affordances and the limitations of publicly available data for each platform. Level-1 engagement is measured on YouTube, as it is the only social media platform that provides viewing of posts. Level-2 and Level-3 engagement are measured on Facebook, Instagram, Twitter, YouTube, and Reddit. Level-4 engagement is calculated for Reddit.

In this research, we seek to understand how user engagement levels across multiple social media platforms are affected by organizational content within the news domain. News organizations have various preferences for news topics that may directly affect how users engage with that content. These news organizations typically operate simultaneously on multiple social media platforms. There is limited prior research involving studies on multiple social platforms (Glenski, Weninger, and Volkova 2018; Vicario et al. 2017) and even fewer ones that focus on a large number of content producers within a single domain (Rieis et al. 2015).

With findings from this research, one can build models or systems that enhance user engagement at the different levels in our framework while the content is being crafted for publication. This aim is the motivation of our research questions, which are:

- RQ1: (a) Do content topics differ among social media platforms? (b) Do content topics among social media platforms differ by organization?
- RQ2: (a) Is user engagement affected by the content topic? (b) If so, which topics generate higher levels (e.g., Level 4 is higher than Level 1) and volume of user engagement?
- RQ3: Can we predict the content that will receive the highest volume of Level-4 user engagement?

This research is novel in several regards. First, rather than focusing on a single metric, we employ a spectrum of metrics and present these various metrics within an overall framework of levels of user engagement based on the degree of public expressiveness, which is both original and provides needed order to the array of social media metrics. Second, most prior research has been conducted on a single platform, often within the confines of a single organization or a single event. In contrast, our research conducts a multi- and cross-platform analysis of four social media platforms (Facebook, Instagram, Twitter, and YouTube) and a community-based social network (Reddit). Along with the multiple platforms, our research entails 53 organizations within the news industry during an 8-month data collection period for posts that cover a variety of topics. Finally, the limited prior work that has been done in cross-platform analysis has typically focused on individual users. Instead, we focus on the organizational level, the content producers and their use of multiple social media platforms to disseminate content, and how individuals interact with this content and share it on other public networks. As such, our research has the potential to substantially impact our understanding of the social media strategies of content producers who engage in the cross-platform dissemination of content and the effect of user engagement within this context.

RELATED WORK

User Engagement Modeling

User engagement is a key concept for many web applications (Lehmann et al. 2012), including online content, which users can interact with in various ways across multiple platforms. While user engagement encompasses various perspectives, the approach that we take for this research is an analytical approach in which user engagement is a recorded user behavior (Jansen 2009; Lalmas, O'Brien, and Yom-Tov 2014; An and Weber 2018); for this study, that means specifically interacting with online news content posted on social media platforms. As such, we are focused on behavior within social media analytics that is used to measure user engagement. Other researchers have employed aspects of these user analytic metrics in prior work. For example, Zhang et al. (2017) uses log files to collect user behaviors to predict how engaged users are with a task during searching.

Researchers report that time and query features best predict user engagement. Bhamidipati, Kant, and Mishra (2017) leverage user engagement to predict the probability of users clicking on ads and installing apps. Lagun and Lalmas (2016) examine metrics for user engagement of news article reading, reporting that a small set of metrics can predict whether a user will bounce or read the complete article. Our approach to user engagement builds from these prior works but takes the perspective of user engagement being in a framework based on the level of public expression of involvement with the online content.

Expression on social media posts has been studied for various platforms, such as Facebook (Van Canneyt et al. 2018; Srinivasan et al. 2013; An, Quercia, and Crowcroft 2014), Instagram (Jaakonmäki, Müller, and vom Brocke 2017; Ferrara, Interdonato, and Tagarelli 2014), Twitter (Van Canneyt et al. 2018; Bandari, Asur, and Huberman 2012; An et al. 2014), YouTube (Ma, Yan, and Chen 2017; Vallet et al. 2015), and Reddit (Stoddard 2015). However, the previous work has mostly focused on views, likes, shares, and comments on the same platform, with little focus on how these engagement metrics compare across platforms and how public the user engagement is. In this research, we investigate Level-4 user engagement with news content, measured by the user behavior of posting news articles on a public network. Moreover, we provide a comprehensive view of engagement across different social media platforms from the same set of content providers.

Content Modeling

There are a variety of content features, with one of the most important being topic. Users tend to have different topic preferences on social media (Ferrara, Interdonato, and Tagarelli 2014; Guo et al. 2015), and the content topics have been used in prior research to predict user preferences, such as product recommendations (Zhang and Pennacchiotti 2013). On Twitter, topics have been used to understand user preferences in order to generate higher volumes of engagement (Yang and Rim 2014). Understanding of individual user topic preferences has been used to filter news posts for individual users on social media platforms, such as Facebook and Twitter (Kazai, Yusof, and Clarke 2016). Researchers have also used topic analysis to detect events from vast streams of breaking news and evolving news story structures (Altınel and Ganiz 2018). For example, researchers can leverage 60 GB of real-world news data to accurately identify events (Liu et al. 2017). Zarrinkalam, Kahani, and Bagheri (2018), combine topic classifications, and predict those users who will share the content. Most prior work on topic classification has studied only one social media platform, with limited studies using multiple platforms (Lee, Hoang, and Lim 2017; Mukherjee and Jansen 2017). We leverage topic analysis in our research to determine organization topic preferences on social media platforms and levels of user engagement based on content topics across multiple social media platforms.

To identify topics, prior work has leveraged Latent Dirichlet Allocation (LDA), which is a statistical model that is widely used in textual analysis (Blei, Ng, and Jordan 2003). LDA identifies hidden topics (e.g., sports, education) from large collections of text articles, where each article is seen as a set of topics with different distribution. However, applying standard LDA on short text (e.g., tweets, social media posts) has proven ineffective, as each short text message most likely represents one or a limited topic set (Jelodar et al. 2017), so, the probability of being within a topic is nearly binary. Twitter-LDA was proposed to overcome this short text problem (Zhao et al. 2011). Fang et al. (2016) reported that Twitter-LDA captures more meaningful topics than original LDA for short-text documents. Social media applications have adopted Twitter-LDA to get meaningful topics from micro-blogs rather than employing the standard LDA (e.g., bursty topic detection (Diao et al. 2012), aspect mining (Yang, Chen, and Bao 2016), and user modeling (Jiang, Qiu, and Zhu 2013)). Based on these advantages, we adopt Twitter-LDA for our research.

Data Collection

To investigate our research objectives, we develop a list of (a) news organizations and (b) the social media platforms that these organizations post content to. We then collect the content that these news organizations posted to the social media platforms, along with the audience engagement numbers for each post.

Selection of Media Organizations

Considering five different rankings of news sources, including PewResearch[1] and Wallethub[2], we construct a list of 60 news organizations that are the top English-based international online sites. We then examine the social media channels for each of the news organizations to eliminate the inactive ones, resulting in the exclusion of 7 news organizations. We use the remaining 53 news organizations for this research. The 53 final news organizations are based in five different counties, with 44 being US-based, 5 being UK-based, and 4 being based in other countries. We do not list the 53 news organizations due to space limitations, but some of the organizations, as examples, are BBC, CNN, Fox, The New York Times, The Economist, TIME, BuzzFeed, VICE, ABC, Vox, and Bleacher Report.

Selection of Social Media Platforms

For this multi- and cross-platform social media research, we select the four most popular social media platforms used

[1]http://www.journalism.org/2011/05/09/top-25/

[2]https://wallethub.com/blog/best-news-sites/21699/

most by the news organizations for posting content, which are Facebook, Instagram, Twitter, and YouTube (Kallas 2017). We then identify the verified social media accounts for each organization on each of these platforms. For each platform, we select the one social media account that focuses on general news posts for each news organization to ensure our dataset covers diverse set of topics for our analysis. For example, the BBC has many Twitter accounts; however, we select BBCNews, as it contains general news postings.

Also, we collect content from Reddit, which is one of the most popular online communities. Unlike the other social media platforms in this study, the news organizations do not have official accounts on Reddit. However, users publicly share and discuss news articles across various Reddit sub-communities.

Data Description

The collection period spanned over eight months, January through August 2017, inclusive. The number of collected posts, likes, and comments from each social media platform for the 53 organizations are shown in Table 2. All 53 news organizations are active across the four social media platforms with the exception of two organizations that are inactive on YouTube. Also, the Associated Press and Mic have disabled the comments feature for their YouTube videos

Platform	# Media	# Posts	# Comments	# Likes
Facebook	53	27,117	984,266	70,557,281
Instagram	53	35,289	11,732,837	723,493,279
Twitter	53	571,270	14,426,570	13,604,785
YouTube	51	43,103	4,674,630	33,265,610
Reddit	53	2,486,594	18,200,179	147,521,797
Total		3,163,373	50,018,482	988,442,752

Table 2: Summary of social media posts and associated user interactions for the 53 news organizations

Facebook: We build a web crawler to collect content from the 53 selected news organizations pages via the Facebook API, collecting all posts within the 8-month period. With each post, we also collect the associated user engagement metrics (see Table 2).

Instagram: We implement an Instagram crawler taking the account name of the news organization as input and returning all Instagram posts, metadata, and user engagement metrics. We then filter the posts to retain only posts published within the eight-month period, again, collecting associated engagement metrics (see Table 2).

Twitter: The tweets of each organization were collected using a web crawler that returned all tweets IDs for each organization within the specified time range. We then used the Twitter API to collect the tweets and user engagement metrics (see Table 2). The use of this method overcomes any potential bias inherent in the Twitter Streaming API, as we are collecting all posts from a news organization as posted.

YouTube: We use the search function of the YouTube Data API to collect the list of video IDs on the channel of each news organization, along with the video engagement metrics (see Table 2).

Reddit: We use a publicly-available Reddit dataset[3]. We extract the submissions using the news organizations websites' domain names from January until August 2017. As a result, we extract 2,486,594 Reddit posts with associated metadata, such as scores[4] and the number of comments (see Table 2). However, we observe that a lot of Reddit posts were produced by automated Reddit accounts. Thus, we need to distinguish between articles posted by bots and users. We assume that users with high activity levels are more likely to be bots. Therefore, we manually examine the top 100 users ranked by posting volume to identify the common patterns of bots on news postings. We eliminate those who posted more than 1,000 posts and had one of the following words in their username: bot, auto, news, or admin. After applying this filtering, we remove 796 users[5] out of 128,956 unique users and their posts in our dataset. As a result, we have 128,160 unique Reddit users and their 602,870 posts.

Methodology

Engagement Metrics

We calculate four engagement metrics, which are: (a) Normalized View (NV), (b) Normalized Likes (NL), (c) Normalized Comments (NC), and (d) Normalized External Posting (NEP). Since those values are highly skewed, we normalized the number of views, likes, and comments using the log-normalization, adding one to each number to account for zero values. Likes and comments are common and measurable across all the five platforms; also, Reddit has a score that is a function of up and down votes that we refer to as likes for commonality in terminology across platforms. NV is only accessible on YouTube, and NEP is only accessible on Reddit. To compute NEP, we first calculate external posting count (EPC) for each news article, which is how many times the same article is posted on the public network. Then, we calculate NEP using the log-normalization similar to the other engagement metrics. Although in our research, we examine external posting behavior only on Reddit, one could also examine this metric on other public platforms where data access is available.

We examined the normalization in two ways by dividing by the number of followers for each news media and without doing so. We found the results of the topic analysis were similar in both cases. We note that we did not consider normalization by the number of days from post time to collection time because most engagement activities happen during the first couple of hours from posting time. Hence, day normalization would give unfair engagement metrics values to posts. For example, a tweet with 1000 likes published 60 days before collection time would have a lower engagement metric (1000/60=16), while a tweet with similar likes (1000) published 30 days before collection time would have

[3]https://files.pushshift.io/reddit

[4]On Reddit, people can do an up-vote or down-vote for a post, and the top score is computed by the number of up-votes minus down-votes.

[5]426 usernames include 'bot', 62 include 'auto', 36 'admin', and 203 include 'news'.

a higher engagement metric (1000/30=33). We also calculated the normalized share (NS) (e.g., sharing on Facebook or retweeting on Twitter), but we later found it was highly correlated with NL, so we did not report that here.

Engagement Levels

Although some user engagement metrics are common across most social media platforms, there are other metrics that are unique to specific platforms. This highlights the benefit of studying multiple social media platforms. A comprehensive study of organizational presence across multiple platforms makes the findings clearer and more complete as there are different types and levels, from private to public, of engagement on different platforms, as presented in Table 1.

Due to the affordances of each platform and the limitations of data collection, different user engagement metrics are accessible for each platform. Level-1 engagement can be measured on YouTube, as it is the only social media platform API that returns viewing of posts. Level-2 and Level-3 engagement can be measured on Facebook, Instagram, Twitter, YouTube, and Reddit. Level-4 can be measured for Reddit.

Content Analysis

In order to understand how the postings and user engagements differ among platforms and engagement levels, we conduct a text-based topic analysis.

Topic Modeling with Twitter-LDA: We build a topic model for the news posts on social media platforms using Twitter-LDA (Zhao et al. 2011), which is a variant of LDA designed for short social media posts.

Data Cleaning: To build a topic model, we adopt the following data-cleaning steps for each of the social media posts in our dataset. First, we remove all URLs, email addresses, and punctuation. We further remove stop words, domain-specific terms (e.g., news, article, etc.), and news organizations' names (e.g., BBC, CNN, etc.). We then apply tokenization and stemming (Porter Stemmer). Finally, the posts with less than five words are deleted.

Determining the Number of Topics: For topic analysis, we need to determine the number of topics. To select the optimal number of topics, we examine the coherence among topics of a model (Fang et al. 2016). The level of coherence is calculated based on the top n most probable words for each topic t. We construct all pairs of the top-n words. Then, for each pair, we compute pointwise mutual information (PMI), a well-known method for semantic similarity (Fang et al. 2016). The topic coherence (C) is the average PMI score across all word pairs. More formally, the C is defined as follow:

$$C(t) = \frac{1}{\sum_{n-1}^{m=1} m} \sum_{i=1}^{n} \sum_{j=i+1}^{n} PMI(w_i, w_j) \tag{1}$$

, where n is the number of top words, and the PMI score of a word pair (w_i, w_j) is computed as follow:

$$PMI(w_i, w_j) = log \frac{p(w_i, w_j)}{p(w_i) * p(w_j)} \tag{2}$$

, where $p(w_i)$ is a probability of observing word w_i, and $p(w_i, w_j)$ is a probability of the two words w_i, w_j appearing in the same post.

A higher value of the average PMI indicates better coherence. We examine 10 different Twitter-LDA models built by a different number of topics (k), from 10 to 100, at a different number of top words (n) @(5, 10, 20). Figure 1 shows the average PMI at different top n words over an increasing number of topics. We observe that the average PMI is the highest when $k = 20$ for all three n values with the highest PMI scores are equal to 3.5 for coherence@5. As a result, we use Twitter-LDA with 20 topics for all the analyses in this study. Once we have the topic model, we assign each of the social media posts to one of the 20 topics. Also, for each of the 20 topics, we manually examine the top 20 words to assign a human-readable topic label.

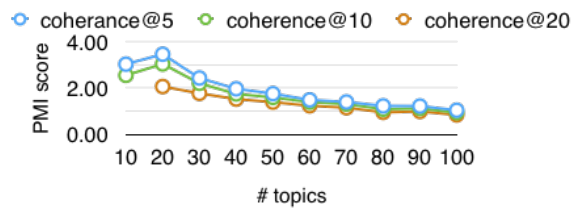

Figure 1: The coherence of the Twitter-LDA model with different number of topics k at four coherence average points

Building Prediction Models for External Posting Behaviors

While previous work studied likes, comments, etc. at Levels 1 to 3 (Van Canneyt et al. 2018; Jaakonmäki, Müller, and vom Brocke 2017; Vallet et al. 2015), Level-4 (public posting) engagement is less-studied. Thus, in this research, we focus on external posting behaviors, and we build two prediction models.

First, we predict whether a news article would be publicly posted, specifically on Reddit. Our data includes only the 'positive' cases for this prediction task. Thus, we need to create 'negative' cases which are news articles that have not been posted on Reddit. To this end, we collect all news URLs using news organization RSS feeds during our data collection period. We then download the news articles for all valid URLs, resulting in 914,671 unique news articles. Then we remove all news articles shared on Reddit, resulting in 535,841 (58%) articles. After that, we randomly sample news articles from those remaining in order to generate a balanced dataset. Our model's random baseline is 0.5.

Second, we build a model that predicts whether a news article would be posted publicly multiple times on Reddit or not. The number of times a news article is posted on Reddit is a strong indication of a high Level-4 engagement among multiple users. An article publicly shared once is different from an article published more than one time, as the article will go into different Reddit communities, which increases the public share of the article. To construct training data for this model, we consider news articles posted twice or more as 'positive' cases and those posted once as 'negative' cases.

We sub-sample the negative cases to make the dataset balanced because most articles were posted only once. The random baseline for this task is 0.5, as well.

We note that both models target an application for news producers, and thus, we use the news headlines as an input of the prediction models. For both models, we use different language features that can be grouped into four different categories:

- Topic: the topic of a post assigned by Twitter-LDA model, represented by a one-hot vector of size 20
- Language: a Term Frequency–Inverse Document Frequency (TF-IDF) matrix with maximum IDF of 0.8, minimum IDF of 0.01, and a maximum of 5,000 features (i.e., words)
- Textual: the number of characters, whether a post includes emoji, question mark, or exclamation mark
- Sentiment: a sentiment label (negative, neutral, or positive) classified by VADER (Valence Aware Dictionary and sentiment Reasoner), which is optimized for short texts (Hutto and Gilbert 2014)

As a measure of prediction results, we use Precision, Recall, F1-score, and Area Under the Curve (AUC) with a 10-fold cross-validation. However, since we construct balanced datasets for the two prediction tasks, there are not many differences among the four measures. Thus, we report the F1-score. We test three different classification algorithms: AdaBoost, Decision Tree, and Random Forest.

Results

Exploratory Analysis

To examine content differences across social media platforms and news organizations, we assign each of the social media posts exclusively to one topic using the Twitter-LDA model (Zhao et al. 2011). Figure 2 shows the number of posts for each of the 20 topics across all platforms. The most popular topic is "Business". In total, there are 264,905 posts on that topic in our dataset. The average number of posts per topic is 157,275, indicating that each topic has enough posts in order to study the differences across platforms and among news organizations.

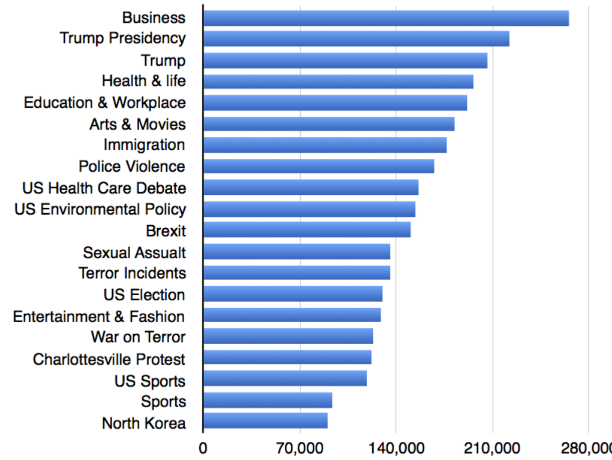

Figure 2: Number of posts assigned to each of the topics across all five social media platforms

Topic Analysis of Platforms

We present here the analysis of the differences and similarities of topic distribution across the social media platforms and discuss the results of how each of the 53 news organizations is posting topics across these platforms.

Platform Popular Topics Turning to RQ1 (a) *Do content topics differ among social media platforms?*, Figure 3 shows the top five common topics for each of the social media platforms. The authors manually assigned the labels of LDA topics by examining the most frequently occurring words within each topic. It appears that news organizations are targeting certain topics for specific platforms. For example, Facebook has education and workplace as the most common topics (13%). However, Instagram has 17% of posts in the health and life category, with entertainment being the second most common topic (15%). Twitter, YouTube, and Reddit have business as their most prevalent topic (13%, 9%, and 7%, respectively).

From the pie charts in Figure 3, we notice both similarities and differences among the five platforms. A few of the topics have universal or near-universal appeal that is independent of the social media platform. Some of the top five topics on each platform are common; there are 11 unique topics. Interestingly, Health and Life are popular on all five platforms, and the topics of Business and Trump are popular on four of the five platforms. Conversely, there are five topics (e.g., Arts, Sports, etc.) that are popular on just one of the platforms. So, it is apparent that there are some platforms, either through affordances, type of medium, or make-up of users, that lend themselves to the dissemination of certain topics.

The differences in topic distribution among platforms indicate that news organizations are, in some cases and for some topics, using social media platforms for different content-dissemination purposes. This disparity would indicate that organizations are responding to each platform's unique features, which may lend themselves to the dissemination of given types of content, and also responding, perhaps, to their users' unique preferences.

Organization Topic Similarity across Platforms Moving to RQ1 (b) *Do content topics among social media platforms differ by organization?*, in analyzing the degree of similarities of topics across platforms at the organizational level, we use the Jensen-Shannon Divergence (JSD) score. JSD is a method for measuring the degree of similarity between two probability distributions (Lin 1991). As the posts on Reddit are not directly from the news organizations, we do not consider them in this analysis. Therefore, we present the topic similarities among the remaining four platforms. For the six platform pairs (e.g., Facebook-Twitter, Facebook-Instagram, etc.), we compute the JSD score for a

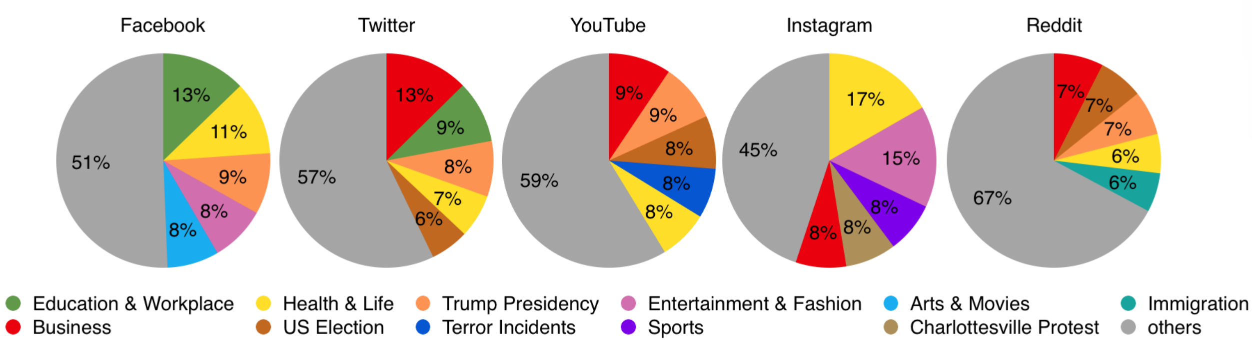

Figure 3: Pie charts for top five topics on each social media platform

news organization (org.) on any platform pair p_i and p_j as follows:

$$JSD(p_i||p_j|org.) = \frac{1}{2}D(p_i||p_j|org.) + \frac{1}{2}D(p_j||p_i|org.) \tag{3}$$

, where JSD is based on Kullback-Leibler divergence $D(p_i||p_j|org.)$, which is calculated using the following equation:

$$D(p_i||p_j|org.) = \sum_k P(k|p_i, org.) \log \frac{P(k|p_i, org.)}{P(k|p_j, org.)} \tag{4}$$

, where k is number of topics ($k = 20$ in our case) and $p(k|p_i, org.)$ is the probability of topic k when organization org. posts on platform p_i.

By measuring the JSD score, we know to what extent each news organization is sharing similar or dissimilar topics across Facebook, Instagram, Twitter, and YouTube. The JSD score value ranges from 0 to 1, where 0 means completely dissimilar topics between platforms within each pair and 1 means the news organizations distribute identical topics between platforms. For example, if the JSD score for NYTimes on Facebook and Twitter is 1, then the topic distribution of the NYTimes posts on the two platforms are the same. Figure 4 shows the distribution of the number of news organizations with given JSD scores for the six social media platform pairs. Of all platform pairs, Twitter-Facebook has the lowest average JSD score across 46 news organizations. We find that most organizations have low JSD scores, indicating that they do not share similar topics across platforms. Of the 306 pairs[6], 141 (46%) have a JSD score lower than 0.1. However, there are seven (7) organizations with JSD scores greater than 0.5, meaning they share similar topics between YouTube-Instagram, Twitter-Instagram, and Instagram-Facebook. The overall low similarities of content across social media platforms for individual news organizations, as shown by the JSD scores, indicate that a model predicting user engagement on one platform is likely not to be transferable to another platform.

[6]We used 51 news organizations for this analysis because the number of available news organizations for YouTube is 51.

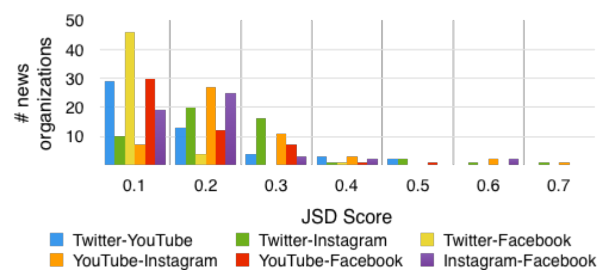

Figure 4: JSD scores for the six social media platform pairs.

User Engagement and Topical Effect

Now we know that news organizations are posting topics differently across platforms, we study user engagement for each of the 20 topics across the five platforms for the 53 news organizations. We also include Reddit for this analysis (i.e., Level-4 engagement). This analysis addresses both RQ2 (a) *Is user engagement affected by the content topic?* and RQ2 (b) *If so, which topics generate higher levels (e.g., Level 4 is higher than Level 1) and volume of user engagement?*.

First, we divided the posts within each topic into three equal portions using a 33% quantile based on each selected engagement metric. Then, we consider the top 33% of posts as high-engagement posts (High) and the bottom 33% of posts as low-engagement posts (Low). The middle 33% of posts are ignored to ensure the differences between engagement metric value are comparable. We then calculate the chi-square measure to determine whether the difference between the number of high- or low-engagement posts for each of the 20 topics is statistically significant. We use the chi-square test commonly used for categorical data, which is the topics in our case. We test this for each platform and for NV, NL, NC, NEP, and EPC. For EPC, we compare Reddit posts posted once versus those posted multiple times.

The chi-square significant test results for selected topics are shown in Table 3. It is apparent that user engagement is significantly different among topics, and user engagement varies by topic from one platform to another. We see two general classes of topics. First, there are topics (e.g., Charlottesville Protest) that have significantly high engagement across most platforms or generally have significantly low engagement across most platforms (e.g., Immigration). Second, there are topics that have mixed engagement, both high and low, depending on the specific platform (e.g., Business – low engagement on YouTube and high engagement on Twitter). Compared to what news organizations publish on social media platforms (Figure 3), we observe that topics that are published more by news organizations are not necessarily the ones users engage with more. For example, on Facebook, education is the topic with the most posts by news media; however, it has no significant correlation with engagement.

Furthermore, there are certain topics with different user engagement at different levels. On YouTube, Terror Incidents-related posts tend to receive many views but not many likes and comments. Across two platforms (Instagram, YouTube), the topic Immigration generally has low NL engagement but high NC engagement on YouTube. These results can help news producers target a particular platform for a particular level of engagement given the news content. For example, if one wants to know what people think about a movie, then Twitter could be the best choice, as arts and movies is the topic that Twitter users actively comment on. The highest user engagement level varies based on the platform. For Twitter and Reddit, NL and NC show similar patterns. On YouTube, one topic (Business) shows low engagement across three levels, NV, NL, and NC.

In addition, we calculated the chi-square of content posted only once versus multiple times, denoted as EPC, as shown in Table 3. The results show that different topics have multiple postings, which indicates high engagement in Level-4 (e.g. Entertainment and Arts, Movies). The result implies that for some platforms, the expected user engagement behaviors are similar across different engagement levels, but for other platforms, user engagement levels differ significantly. News organizations must identify their key performance indicators and adjust their topical posting strategies accordingly per platform.

Predicting Level-4 Engagement

We now move to RQ3 (*Can we predict the content that will receive the highest volume of Level-4 user engagement?*). In particular, we build two prediction models: 1) whether a news article will be posted on Reddit by users, and 2) whether a news article will be posted multiple times on Reddit by users.

Predicting Whether a News Article Will Be Posted on Reddit by Users We build an external posting prediction model for Level-4 engagement for each news organization. To this end, we construct a balanced dataset that includes 2,000 random samples of positive cases (news articles posted on Reddit) and 2,000 random samples of negative cases (news articles not posted on Reddit) for an individual news organization. Some news organizations do not have enough data, and thus we exclude them, resulting in 29 news organizations for this experiment. We extract four sets of features: topic, language, textual, and sentiment from those sampled posts and group them into two super-sets: language features and metadata features. Metadata features include topic, textual, and sentiment features. With these input features, we train the model using three different algorithms (i.e., AdaBoost, Decision Tree, and Random Forest) for each of 29 news organizations. For language features, we conduct a random forest feature selection to reduce the number of features to 1,000 to prevent over-fitting. We run different experiments to tune the parameters of each machine learning algorithm and report the best results from the tuned model. Table 4 reports the average, minimum, maximum, median, and standard deviation of the F1-score[7].

When using language features, a decision tree results in the best performance with an F1-score of 0.68 while a random forest works better with metadata features with an F1-score of 0.67. Overall, the models trained by the decision tree that use all the features outperform the two other models with an average F1-score of 0.71. However, the prediction performances significantly vary among different news organizations. For example, the maximum F1-score is 0.93 for the Forbes news organization, but the minimum F1-score is 0.62 for BBC News for the model built with the decision tree with all the features. Our results show that features extracted from posts can predict Level-4 user engagement with 70% precision, on average. Therefore, our model can help news producers to know whether a news article will be publicly shared before publishing the article on some other platform. In particular, even with simple metadata features, which include topic, textual, and sentiment features, our model achieves 68% precision on average in predicting Level-4 user engagement.

Predicting Whether a News Article Will Be Posted Multiple times on Reddit by Users In the second prediction task, we build a model that predicts whether a news article will be posted multiple times or just once. Intuitively, an article posted multiple times on Reddit has greater exposure than an article posted only once, given the sub-community structure of Reddit. We first construct a balanced dataset for this prediction task. Since few news articles are posted multiple times on Reddit, there are far fewer positive cases. We consider news organizations that have at least 2000 positive cases (posted multiple times) and 2000 negative cases (posted only once), resulting in using 29 news organizations. Similar to the previous prediction tasks, we evaluate the three algorithms (i.e., AdaBoost, Decision Tree, and Random Forest) with two sets of features (i.e., language and metadata).

We report the best prediction results by the model based on Decision Tree with all features due to space limitations. Table 5 shows the prediction results by F1-score. Our model predicts at 85% precision and F1-score of 0.83 on average those articles that are posted multiple times on Reddit. We find that this prediction task performs better than the first prediction task. The result indicates that predicting which news article will be posted on Reddit (Level-4 engagement) is more difficult than predicting which news articles will be posted multiple times on Reddit. By combining the two

[7] We omit the precision (P), recall (R), and area under the curve (AUC) as those measures are very similar to the F1-score given that our data is balanced.

Platform	YouTube			Facebook		Twitter		Instagram		Reddit			
Level	1	2	3	2	3	2	3	2	3	2	3	4	4
Eng. Metric	NV	NL	NC	NL	NC	NL	NC	NL	NC	NL	NC	EPC	NEP
Business	Low***	Low***	Low***			High***	High***				High***	High***	
Trump Presidency	High***	High***	High***	High*			High***		High***	High***	High***	High**	
Health & life	Low***		Low***				High***	High**	Low**	High*	High***	High***	Low*
Education	Low***		Low***			High*				High*	High**	High***	
Arts, Movies	Low*		Low***			High***	High***				High***	High***	
Immigration	Low**	Low**	High***					Low**		High**			
US election	Low***	Low***				High***	High**				High***		
Terror Incidents	High***	Low***	Low***		Low**		High***				High***		
Entertainment	Low***		Low***			High***	High***			High*	High***	High***	
Sports			Low**				High***			High***	High***		
Charlottesville	High***	High***	High***				High**		High***	High***	High***		

Table 3: Chi-square results for selected number of topics. Low-engagement and high-engagement topics are represented as Low or High, respectively. For EPC, we compare Reddit articles posted once versus those posted multiple times.

		F1-score				
	Features	Avg.	Min	Med	Max	Std
AdaBoost	L	0.62	0.54	0.61	0.93	0.09
	M	0.63	0.54	0.63	0.78	0.06
	L+M	0.68	0.58	0.67	0.94	0.08
Decision Tree	L	0.68	0.57	0.66	0.93	0.08
	M	0.66	0.56	0.66	0.86	0.07
	L+M	0.71	0.62	0.71	0.93	0.08
Random Forest	L	0.64	0.55	0.62	0.93	0.08
	M	0.67	0.57	0.65	0.80	0.06
	L+M	0.67	0.56	0.66	0.94	0.08

Table 4: The prediction results for three features sets (languages (L), metadata (M), and (L+M)). average (Avg.), minimum (Min), median (Med), maximum (Max), and standard deviation (Std.) for F1-scores for 29 news organizations.

models, we can predict whether a news article will have high engagement on Reddit (Level-4) before publishing the news article, which addresses RQ3.

	Avg.	Min	Med	Max	Std
F1-score	0.83	0.63	0.85	0.94	0.08

Table 5: The F1-score by one set feature (languages (L)+metadata (M)) by decision tree algorithm across 29 news organizations.

Discussion and Implications

Measuring user engagement with content on social media platforms is a challenging problem to tackle, as there are often different media objectives. It is difficult to identify the user segment to target (i.e., the people to whom one wants to distribute content). It is a challenge to communicate a message that resonates with the spectrum of users. Addressing these challenges often requires a multi-channel approach, which is why we focused on multiple social media platforms in this research. In leveraging the concept of eWOM, we also focus on multiple levels of user engagement metrics, measuring a person's willingness to express engagement publicly.

The publicity of the news media is essential for higher user outreach and content engagement; hence, higher publicity would help news organizations extend their reach. Our findings show that news organizations differ in their publishing strategy on different social media platforms. We observe news organizations have different topical content distributions, indicating that they have adapted to different user bases on each of the social platforms to some degree.

Also, content having high- or low-user engagement differs among platforms, and these differences occur at different engagement levels. Such differences between platforms and engagement levels indicate that user behaviors regarding the content on one platform are less likely to be observed on another platform. Although, some pairs of platforms do show similar user engagement patterns.

User engagement is affected by different content topics and is composed of different user actions across platforms, so the variances are nuanced. Some topics, like the Charlottesville Protests, generate a high volume of engagement (for Level-3) across nearly all social media platforms. However, some topics, like immigration, generally have low-engagement across most platforms and levels. Most interestingly, specific user engagement actions at different levels are associated with different topics. For some platforms, like Twitter or Reddit, the differences between Level-2 and Level-3 engagements are relatively small. However, for other platforms, like Facebook and YouTube, users display different engagement behaviors based on content preferences at Level-2 and Level-3. This may indicate that users are more (or less) willing to show their engagement publicly depending on the topic. The results can help content creators target platforms and engagement levels given their content using a similar analysis as employed in this work. We specifically examined Level-4 engagement. This amplification of content is crucial for user acquisition, as these public sharing actions increase content's reach beyond an organization's existing users. As such, Level-4 user engagement directly or indirectly contributes to user reach.

The prediction power of cross-posting varies among news organizations. One possible basis for this variation may be the different countries in which these news organizations are based. Many are from the US, but some are based in other countries. Reddit is a US-based social media platform, where users mostly share content of US-based news organizations. Hence, this makes it difficult to predict the posting of non-US news organizations using only this platform. Future work can consider predicting external posts on other social media platforms (e.g., Facebook, Instagram, and Twitter), although data access is a hindrance. Moreover, varying levels of global platform adoption may also influence the prediction results, since different nationalities have different usage levels for individual social media platforms.

In terms of strengths, this research is novel in several aspects. First, we focus on multiple organizations within a single domain over an 8-month period across multiple social media platforms. Second, we specifically include Reddit posts by users who spread content from these news organizations. Third, we present the analysis within a four-level framework of user engagement. As such, the research has implications for both practice and theory. This research is one of a small but growing area of cross-platform analysis. It is one of the few studies of multiple organizations within a single domain, and it presents an organized framework for user engagement metrics.

Many future directions are opened by this research. First, we focused on the content features of the article; however, other features can be considered, such as multimedia content (e.g., images and videos) or context (e.g., time and location), the attributes of an article's creator (e.g., age, gender, and number of followers), and the attributes of the social media users who reposted an article. Second, we could expand the level of engagement to include more nuanced metrics such as behaviors prior to and after users' actual postings (Grinberg et al. 2016). Third, an investigation of the private sharing of news content would be interesting. Finally, an insightful area of research would be level-four engagement (i.e., public sharing) on other channels.

Conclusions

We developed a four-level framework for user engagement based on the degree of public expression from more private to more public. We then evaluated user engagement with online content from 53 news organizations across five social media platforms: Facebook, Instagram, Twitter, YouTube, and Reddit. We used 3,163,373 social media postings from these news organizations across the five platforms during an eight–month period with associated engagement metrics. We reported differences in platform topical postings by an organization for each platform and the effect on user engagement by level. Our findings show that topic distribution varies by platform and that user engagement also varies by topic and platform. This research could aid media organizations in understanding how topics affect user engagement at different levels across different platforms. Findings contribute to making better management decisions during content creation.

References

Aldous, K. K.; An, J.; and Jansen, B. J. 2019. The challenges of creating engaging content : Results from a focus group study of a popular news media organization. In *ACM CHI*.

Altınel, B., and Ganiz, M. C. 2018. Semantic text classification: A survey of past and recent advances. *Information Processing & Management* 54:1129 – 1153.

An, J., and Weber, I. 2018. Diversity in online advertising: A case study of 69 brands on social media. In *SocInfo*.

An, J.; Quercia, D.; Cha, M.; Gummadi, K.; and Crowcroft, J. 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science* 3(1):12.

An, J.; Quercia, D.; and Crowcroft, J. 2014. Partisan sharing: Facebook evidence and societal consequences. In *COSN*.

Balbi, S.; Misuraca, M.; and Scepi, G. 2018. Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management* 54:674 – 685.

Bandari, R.; Asur, S.; and Huberman, B. A. 2012. The pulse of news in social media: Forecasting popularity. In *ICWSM*.

Bhamidipati, N.; Kant, R.; and Mishra, S. 2017. A large scale prediction engine for app install clicks and conversions. In *CIKM*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning research* 3:993–1022.

Diao, Q.; Jiang, J.; Zhu, F.; and Lim, E.-P. 2012. Finding bursty topics from microblogs. In *ACL*.

Fang, A.; Macdonald, C.; Ounis, I.; and Habel, P. 2016. Examining the coherence of the top ranked tweet topics. In *SIGIR*.

Ferrara, E.; Interdonato, R.; and Tagarelli, A. 2014. Online popularity and topical interests through the lens of instagram. In *HT*.

Glenski, M.; Weninger, T.; and Volkova, S. 2018. Identifying and understanding user reactions to deceptive and trusted social news sources. *CoRR* abs/1805.12032.

Grinberg, N.; Dow, P. A.; Adamic, L. A.; and Naaman, M. 2016. Changes in engagement before and after posting to facebook. In *CHI*.

Guo, W.; Wu, S.; Wang, L.; and Tan, T. 2015. Social-relational topic model for social networks. In *CIKM*.

Hutto, C. J., and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*.

Jaakonmäki, R.; Müller, O.; and vom Brocke, J. 2017. The impact of content, context, and creator on user engagement in social media marketing. In *HICSS*.

Jansen, B.; Zhang, M.; Sobel, K.; and Chowdury, A. 2009. Twitter power: Tweets as electronic word of mouth. *Journal of the Association for Information Science and Technology* 60:2169–2188.

Jansen, B. 2009. Understanding user-web interactions via web analytics. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1:1–102.

Jelodar, H.; Wang, Y.; Yuan, C.; and Feng, X. 2017. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Computing Research Repository*.

Jiang, J.; Qiu, M.; and Zhu, F. 2013. It is not just what we say, but how we say them: Lda-based behavior-topic model. In *SDM*.

Kallas, P. 2017. Top 15 most popular social networking sites. http://www.ebizmba.com/articles/social-networking-websites. Accessed: 04 January 2019.

Kazai, G.; Yusof, I.; and Clarke, D. 2016. Personalised news and blog recommendations based on user location, facebook and twitter user profiling. In *SIGIR*.

Lagun, D., and Lalmas, M. 2016. Understanding user attention and engagement in online news reading. In *WSDM*.

Lalmas, M.; O'Brien, H.; and Yom-Tov, E. 2014. Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 6:1–132.

Lee, R. K.-W.; Hoang, T.-A.; and Lim, E.-P. 2017. On analyzing user topic-specific platform preferences across multiple social media sites. In *WWW*.

Lehmann, J.; Lalmas, M.; Yom-Tov, E.; and Dupret, G. 2012. Models of user engagement. *UMAP*.

Lin, K.-Y., and Lu, H.-P. 2011. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior* 27:1152 – 1161.

Lin, J. 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory* 37:145–151.

Liu, B.; Niu, D.; Lai, K.; Kong, L.; and Xu, Y. 2017. Growing story forest online from massive breaking news. In *CIKM*.

Ma, C.; Yan, Z.; and Chen, C. W. 2017. Larm: A lifetime aware regression model for predicting youtube video popularity. In *CIKM*.

Mitchell, A.; Holcomb, J.; and Weise, R. 2017. State of the news media 2017. *Pew Research Center*.

Mukherjee, P., and Jansen, B. 2017. Conversing and searching: The causal relationship between social media and web search. *Internet Research* 27:1209–122.

Noguti, V. 2016. Post language and user engagement in online content communities. *European Journal of Marketing* 50:695–723.

Rieis, J. C.; Benevenuto, F.; Pedro; Prates, R. O.; Kwak, H.; and An, J. 2015. Breaking the news: First impressions matter on online news. In *ICWSM*.

Shearer, E., and Gottfried, J. 2017. News use across social media platforms 2017. *Pew Research Center*.

Srinivasan, B. V.; Natarajan, A.; Sinha, R.; Gupta, V.; Revankar, S.; and Ravindran, B. 2013. Will your facebook post be engaging? In *UEO*.

Stoddard, G. 2015. Popularity and quality in social news aggregators: A study of reddit and hacker news. In *WWW*.

Thonet, T.; Cabanac, G.; Boughanem, M.; and Pinel-Sauvagnat, K. 2017. Users are known by the company they keep: Topic models for viewpoint discovery in social networks. In *CIKM*.

Vallet, D.; Berkovsky, S.; Ardon, S.; Mahanti, A.; and Kafaar, M. A. 2015. Characterizing and predicting viral-and-popular video content. In *CIKM*.

Van Canneyt, S.; Leroux, P.; Dhoedt, B.; and Demeester, T. 2018. Modeling and predicting the popularity of online news based on temporal and content-related features. *Multimedia Tools and Applications* 77:1409–1436.

Vicario, M. D.; Gaito, S.; Quattrociocchi, W.; Zignani, M.; and Zollo, F. 2017. Public discourse and news consumption on online social media: A quantitative, cross-platform analysis of the italian referendum. *CoRR* abs/1702.06016.

Yang, M.-C., and Rim, H.-C. 2014. Identifying interesting twitter contents using topical analysis. *Expert Systems with Applications* 41(9):4330–4336.

Yang, Y.; Chen, C.; and Bao, F. S. 2016. Aspect-based helpfulness prediction for online product reviews. In *ICTAI*.

Zarrinkalam, F.; Kahani, M.; and Bagheri, E. 2018. Mining user interests over active topics on social networks. *Information Processing & Management* 54:339 – 357.

Zhang, Y., and Pennacchiotti, M. 2013. Predicting purchase behaviors from social media. In *WWW*.

Zhang, Q.; Ye, T.; Essaidi, M.; Agarwal, S.; Liu, V.; and Loo, B. T. 2017. Predicting startup crowdfunding success through longitudinal social engagement analysis. In *CIKM*.

Zhao, W. X.; Jiang, J.; Weng, J.; He, J.; Lim, E.-P.; Yan, H.; and Li, X. 2011. Comparing twitter and traditional media using topic models. In *ECIR*.