Modeling and Predicting News Consumption on Twitter

Claudia Orellana-Rodriguez
Insight Centre for Data Analytics
School of Computer Science
University College Dublin, Ireland
claudia.orellana@insight-centre.org

Mark T. Keane
Insight Centre for Data Analytics
School of Computer Science
University College Dublin, Ireland
mark.keane@ucd.ie

ABSTRACT

While much is known about how people tweet and interact on Twitter, surprisingly little is known about how the news items tweeted by journalists – *news tweets* – act as a distribution channel for the news that is spread by social media reading and sharing. This paper aims to fill this gap by analyzing the dynamics of news on Twitter, by revealing what drives users to consume news, and by developing a news consumption prediction model. We present the Twitter News Model (TNM), a computational data-driven approach to elucidate the dynamics of news consumption on Twitter. We apply the TNM to a dataset of interactions between users and journalists/newspapers to reveal *what* drives users' consumption of news on Twitter, and predictively relate users' news beliefs, motivations, and attitudes to their consumption of news. Our findings reveal that news motivations, followed by news attitudes and news beliefs, impact users' behavior of news consumption on Twitter.

CCS CONCEPTS

• **Human-centered computing** → **Web-based interaction**;

KEYWORDS

News consumption, digital journalism, social media, audience engagement, news, twitter

ACM Reference Format:
Claudia Orellana-Rodriguez and Mark T. Keane. 2018. Modeling and Predicting News Consumption on Twitter. In *UMAP '18: 26th Conference on User Modeling, Adaptation and Personalization, July 8–11, 2018, Singapore, Singapore.* ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/3209219.3209245

1 INTRODUCTION

Daily newspaper reading and the viewing of national TV news, have been traditionally correlated with a civic obligation to stay informed about current events [21]. For older people, the daily habit of following the news and this civic sense seems to have persisted as news has moved online and onto social media. For younger people, news consumption has, perhaps, become a more personalized activity to gather information about events that directly affect them or, more a matter of social interaction, as they use news items in online conversations with friends and colleagues [1].

For news providers, it is critically important to understand the dynamics of news consumption in this new social media context. They need to understand the beliefs, motivations and attitudes that drive news consumption in social media users. Such insights should allow news providers to better serve their audiences and motivate future news consumption [24]. Indeed, users who read the news on social media often play an important role in modern journalism, not only as direct consumers, but also as gatekeepers, with almost half the social media users sharing and reposting news stories, images, or videos, and discussing news issues or events online [4]. Also, journalists who post/share their news and interact with social media audiences, can use their knowledge of such factors to influence people's satisfaction [17], which may in turn, have a positive impact in news consumption.

Social media is a particular environment in which news consumption is not a passive activity, but rather one in which users interact with news providers, form communities, express their interests, ask questions, or request more information, via mechanisms such as shares, likes/dislikes, retweets, or mentions.

Figure 1 illustrates some of these dynamics of news consumption, as seen in the interactions of an Irish Twitter audience with 200 journalistic accounts. Each graph represents news audiences (nodes in gray), journalists (nodes in color), and their interactions (edges from news audiences to journalists nodes). The larger the node, the more interactions it has received. We have labeled the top-10 journalists/news outlets in each graph, according to the number of interactions received from the audience. In Figure 1a, the edges represent the news audience mentioning journalists in their tweets. Corporate accounts such as *@Independent_ie*, *@rtenews*, *@IrishTimes*, and *@thejournal_ie* receive most of the audience mentions; however, individual accounts including political journalist *@gavreilly* and sports journalist *@MiguelDelaney*, receive considerable attention as well. In terms of retweets (see Figure 1b), corporate accounts *@rtenews* and *@IrishTimes* receive the bulk of them, with political journalists *@gavreilly*, *@Oconnellhugh*, and *@colettebrowne* also being important targets of audience retweets.

Overall, Irish news audiences tend to interact (i.e., mentions+retweets) more with corporate and political journalists (see Figure 1c).

In this paper, we propose the *Twitter News Model (TNM)*, a computational data-driven approach to predict and explain people's consumption of news by analyzing their interactions with journalists and news-tweets. Our *Twitter News Model (TNM)* is based on the Motivational Consumption Model (MCM) [19] previously developed to better understand the consumption of "conventional" news. We apply the TNM to an empirical study of news consumption on Twitter, designed to (i) reveal *what* drives users to consume

UMAP '18, July 8–11, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5589-6/18/07…$15.00
https://doi.org/10.1145/3209219.3209245

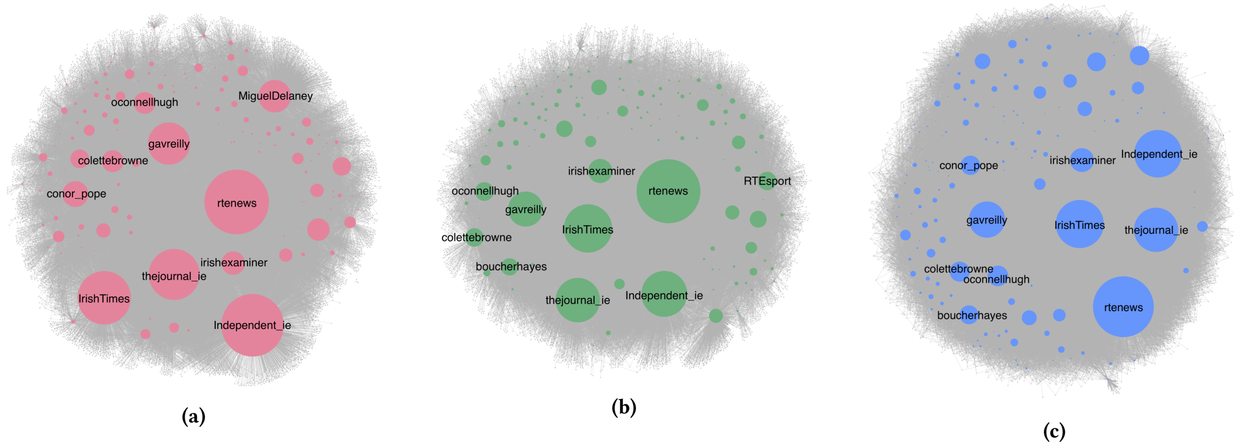

Figure 1: Dynamics of news audiences' interactions with journalists on Twitter. Gray nodes represent the audience and colored nodes represent journalists. Each edge represents a member of the audience (a) mentioning a journalist, (b) retweeting a journalist's tweet, and (c) overall interacting (mentioning + retweeting) with a journalist. The nodes' sizes are proportional to the number of (a) mentions, (b) retweets, and (c) overall interactions they received. On each figure, we have labeled the top-ten most popular Twitter accounts according to the corresponding metric.

news, and (ii) predictively relate users' news beliefs, motivations, and attitudes to their consumption of news.

In the next section we review related work (Section 2), before presenting the theoretical background (Section 3), instantiating our Twitter-specific model for news consumption (Section 4), and predicting news consumption on Twitter (Section 5). We conclude our work and discuss future directions in Section 6.

2 RELATED WORK

Previous work has analyzed the dynamics of news consumption in social media from different angles, be it by exploring the perception of news, by modeling users' engagement with news articles and trust in news providers, or by studying how news is shared.

Perception of news. Users' perception of news may affect their decision on whether or not to read or share news stories. In Reddit, news stories with titles that have been modified by users are more popular (i.e., receive more votes and more comments) than those with their original title [12]. In Twitter and Facebook, which have different demographics of readers, the same news headline may exhibit different patterns of shares [35]. This fact suggests a correlation between demographics and news perception [31]. In Twitter, political news stories that are negative, short, and with high emotional load, tend to receive more shares from the audience [6]. The perception of political news slant can also affect news consumption. Users belonging to a community (i.e. that are connected to each other), tend to consume similar news articles, and these articles have been found to exhibit political slants [18].

Users' engagement with news. The characteristics of a news item and the type of interaction between users and news providers that can signal audience engagement, have also been an ample subject of study. Wu and Chen [37] found that the frequency of interactions between Twitter users and news sources is an indicator of news being perceived as popular, and Keneshloo et. al [15, 16], using social, contextual, and temporal features of Twitter users who share news stories, found that the freshness of a news story is one of the most important factors in predicting its page views. In our previous works [28, 29], we explored whether Twitter users engage with news differently depending on the news category (e.g., sports vs. politics), and found that what attracts people's attention differs considerable. For example, Irish audiences tend to share sports news-tweets when these contain mentions (i.e., *@user*), and the source is a well known journalist. While for business news, the timeliness of news-tweets has a higher impact on engagement.

Trust in news providers. When deciding on whether or not to consume a given news item, the trust that users have in other users, particularly in news providers, may be of great importance. In [5], Choi et al. explored how customers' trust in content is influenced by the content's source. The authors show that User Generated Content has a stronger effect on cognitive trust, than Marketeer Generated Content. De Meo et al. [22, 23] studied how trust relationships can be leveraged to identify those users producing the most helpful content for a community; and that trust is a crucial factor to keep stable levels of user engagement within a group.

News sharing. In online social media, news audiences are active participants of the news distribution cycle, as they can disseminate and spread news to their social networks. In [32], Reis et al. explored the relationship between demographics and news sharing in Twitter. Their findings reveal that male and white users tend to share more news, and that sharing rates vary according to the news category. For example, Asian female audiences are more inclined to share world and health news, while white male audiences focus more on science and technology news. Kalsnes and Larsson [14] analyze the sharing of news articles for four Norwegian news outlets in Facebook and Twitter. They found that people more frequently shared editorial comment pieces about soft news topics, but that depending on the news outlet, this can change, with sensational news and news involving celebrities becoming the center of attention. Bruns et al.'s [3] analysis of Australian news sharing in Twitter, reveals that for many users, the decision on sharing news is highly related to the popularity and credibility of the news source.

News consumption. An important body of research focuses not only on what news is shared or how users engage with news, but on *why* people consume news. Edgerly et al. [8] analyzed news consumption habits in young adults, and found that besides spending fewer total minutes consuming any news than any other age group, young adults nowadays consume less news than young adults 20 years ago. The authors also found that parents highly influence the news consumption habits of their children. Parents' news habits and reinforcement of news consumption can lead to children consuming more news themselves. For many users, news consumption is consequence of incidental finds [2], they get news on their mobile devices because they are constantly connected and not because they look for it. Following ideas from these previous works and, particularly, those introduced by Lee and Chyi in [19], we now present our approach to computationally modeling the psycho-behavioral factors of news consumption on Twitter.

3 MODELING NEWS CONSUMPTION

Our aim in this paper is to model news consumption on Twitter. Our starting point is the prior work that has modeled motivational aspects of human intention and specific models of news consumption in "conventional" news media. Hence, we explore a psychological theory (i.e., the Reasoned Action Model) that has been applied to the consumption of printed and digital news (i.e., Motivational Consumption Model). In this section, we review these previous models before discussing how they can be operationalized to the consumption of news on Twitter (see section 4).

3.1 Reasoned Action Model

Fishbein and Ajzen's Reasoned Action Model (RAM) [9] maintains that there are three main predictors of intention to engage in a behavior, namely attitude, social norms, and self-efficacy. *Attitude* deals with a person's orientation towards performing the behavior. *Social Norms* consider the normative pressure perceived by the person to perform a given behavior in context. *Self Efficacy* relates to the behavioral control the person perceives themselves to have over the target behavior. RAM has been used to understand specific situations where people have manifested intentions to act. For instance, in political science, the RAM has been used to model voters' attitudes towards candidates and political parties, and how these attitudes impact polls and subsequent voting behavior; in public health research, RAM has been used to understand the key beliefs that influence individuals' health care utilization [13].

3.2 Motivational Consumption Model

Lee and Chyi have adapted the RAM framework to understand news consumption in their Motivational Consumption Model (MCM) [19]. In the MCM, a variety of demographic factors are seen as influencing news consumption via the mediators of news beliefs, news motivations, and news attitudes.

This adapted model backgrounds the effects of RAM's social norms and self-efficacy. Social norms play less of a role in news consumption because of the broadly shared belief that it is one's responsibility to follow the news, to stay informed. Self-efficacy also plays a lesser role as people can easily control their own news consumption given the ubiquitous access they have to news content via a variety of to-hand digital devices [19]. In their model, Lee and Chyi go on to describe and operationalize the three key mediators as follows:

– *News beliefs:* refer to the *value* that news has for people, that news can be conceived to be a means to empower and mobilize the public. These news beliefs were operationalized by 7-point Likert scale ratings of survey questions that asked participants (i) how important the news is to you, and (ii) whether being informed was empowering.

– *News motivations:* refer to the *reasons* driving people's consumption of news. For instance, some people consume news because it helps them keep up with current events that are topics of conversations in their social circle or it allows them to make informed decisions in their daily life; whereas others consume news as a source of entertainment. These news motivations were operationalized by combining ratings from five survey questions that asked participants whether they consumed news (i) to find out what is going on in the world, (ii) to keep up with the way your government functions, (iii) to make yourself an informed citizen, (iv) to gain important new information, and (v) to fulfill your need to know.

– *News attitudes:* reflect positive views of the news, that see news consumption as an enjoyable and advantageous behavior. These news attitudes were operationalized by survey questions that asked participants if they agreed with the following statements: (i) getting the news is enjoyable to you, and (ii) getting the news is advantageous to you.

In their survey, Lee and Chyi gathered data from a US-nationwide sample of 1.1K American adults that varied in several demographic features (e.g., gender, age, income, and race). The survey asked people about their news consumption – reading, watching, or listening – on several news sources, including *New York Times*, *CNN*, and *Google News*, in both traditional and digital formats. The responses were analyzed using a regression method. Based on the MCM, demographics such as age and education were found to positively influence news beliefs, and more positive news beliefs were found to lead to better news motivations and news attitudes, which in turn resulted in more frequent news consumption.

In the present paper, we adapt the Motivational Consumption Model to apply it to news consumption on Twitter, by developing new operationalizations of its mediators that were appropriate to this social media context.

4 NEWS CONSUMPTION ON TWITTER

Lee and Chyi applied their Motivational Consumption Model to the consumption of news found in newspapers (i.e., either printed or online), on news aggregators (e.g., *Google News*, *Yahoo News*), and on network Sunday talk shows [19]. The present paper aims to model news consumption on Twitter, which constitutes quite a different context. On Twitter, the dynamics of news consumption differ in that the users do not need to actively buy/access the news but rather browse their *timelines* at any time/place, and find news posts distributed (i.e., tweeted/retweeted) by Twitter accounts from followed journalists or newspapers, or that appear in the timeline because one or more of the other user's followees (e.g., family,

Mediator	Measure	Description
News Beliefs B	$B(u) = \begin{cases} 1, & \text{if } N_u \geq R_u \\ 0, & \text{otherwise} \end{cases}$	How valuable are news for user u?
News Motivations M	$\forall c \in C, M(u,c) = \frac{i_{uc}}{I_u}$	What motivates user u to consume news?
News Attitudes A	$A(u) = \frac{1}{N_u} \sum_{m_u} Polarity(m_u)$	What is user u's approach toward news?
News Consumption C	$C(u) = I_u$	How much does user u consume news in Twitter?

Table 1: New operationalizations for Twitter.

friends, or public figures) has tweeted or retweeted them. In this social media context, our conception of news beliefs, news motivations and news attitudes needs to be adapted with Twitter-specific instantiations of these notions. We will call our adapted model the Twitter News Model (TNM).

4.1 News Beliefs

News beliefs refer to the value that the news has for people (i.e., whether people find news to be important). In Twitter, users express the value that a news-tweet has for them by retweeting it, liking it, or mentioning the author of that tweet.

To quantify the value of news (B) to a Twitter user u, we use the proportion of retweets and mentions resulting from the interaction between the user and news-tweets, journalists, or newspapers. If a user u mentions journalists in her/his tweets more or as many times as she/he retweets news-tweets, then we can assume they value news on Twitter to a greater extent (see Equation 1).

$$B(u) = \begin{cases} 1, & \text{if } N_u \geq R_u \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where N_u is the total number of tweets in which user u has mentioned a journalist or a newspaper, and R_u is the total number of news-tweets that user u has retweeted, in the same period of time.

Mentions, in comparison to retweets, require more effort and time. Mentions help users create awareness of a tweet's author and even spread the author's tweets to new audiences, including non-followers, thus increasing diffusion. By mentioning a journalist or a newspaper, user u might help other users find more tweets from such accounts on a timely manner [36].

4.2 News Motivations

News motivations refer to the reasons that drive people to consume news (e.g., to become an informed citizen). In Twitter, these motivations can be operationalized by tracking how a user interacts with different news categories (e.g., a user who only interacts with tweets about sports may be mainly motivated by entertainment).

To quantify news motivations (M) for news consumption on Twitter, the observed interaction behavior of each user u with tweets/journalists for different news categories is computed as follows:

$$\forall c \in C, M(u,c) = \frac{i_{uc}}{I_u} \tag{2}$$

where C = *{sci & tech, sports, politics, business, breaking news, lifestyle, corporate}* represents the main content categories found for news[1], i_{uc} is the total number of interactions (i.e., retweeting of a news-tweet and/or mentioning a journalist or newspaper) by user u in a particular news category c, and I_u is the total number of interactions (i.e., retweets and mentions) by user u across all news categories. $M(u,c)$ falls in the range of $[0,1]$ where 0 means user u did not interact with category c and 1 means that all interactions of user u were with category c.

It should be noted that using news categories to identify motivations is quite a coarse-grained approach, as precise motivations within a category may vary [19]. A user only interacting with lifestyle news might be motivated by reasons of entertainment or social interaction. A user interacting mostly with business or political news, may be looking to make specific business decisions or just to track broad societal changes. So, while using news categories gives us broad partitions of motivations, a more fine-grained analysis of these motives remains to be discovered. Finally, it should also be noted that the way users interact with the news, is known to vary by news category. For example, Irish Twitter users consuming sports news engage more with news-tweets that contain mentions, while those who engage with science and technology news are influenced more by the temporal arrival of the tweets [29].

4.3 News Attitudes

News attitudes refer to people's overall orientation to the consumption of the news (e.g., whether it is viewed as an enjoyable behavior). In Twitter, these attitudes can be operationalized by tracking the polarity of users' tweets. To quantify the news attitudes (A) of a user u to news consumption on Twitter, the following is computed:

$$A(u) = \frac{1}{N_u} \sum_{m_u} Polarity(m_u) \tag{3}$$

where $Polarity(m_u)$ is the polarity score (i.e., positive or negative) of each tweet in which user u mentions a journalist or newspaper, and N_u is the total number of tweets with mentions – of a journalist or a newspaper – posted by user u.

In our operationalization of news attitudes, we do not consider retweets, only the users' original tweets in which they mention journalists or newspapers. We make this distinction because news attitudes represent a person's tendency to react favorably towards the news [19]. Retweeting a news-tweet may signal users' interests; however, news-tweets are not tailored by the users themselves but rather by the source of the tweet (i.e., a journalist or a newspaper).

[1] A more detailed explanation on how we classify tweets into these categories can be found in Section 5.

Through mentions, a person can express themselves, in their own words, thus giving a better indication of their news attitudes.

4.4 News consumption

News consumption is generally measured by the frequency with which people read/watch/listen to news. In Twitter, we measure a user u's news consumption (C) by the number of interactions between the user and news-tweets, journalists and newspapers, which are defined as follows:

$$C(u) = I_u \tag{4}$$

$$I_u = N_u + R_u \tag{5}$$

where I_u is the number of interactions of user u, N_u is the total number of tweets in which user u mentions a journalist or newspaper, and R_u is the total number of news-tweets that user u retweeted.

We consider that both retweets and mentions give us an indication of the users' overall tendency to consume news. For example, by retweeting a news-tweet, a person may indicate interest in a topic or a specific news article, while by mentioning a journalist or a newspaper, a person may express an interest in the work of the journalist or the coverage of the newspaper. Therefore, we use these interactions as a proxy for news consumption.

In conclusion, Table 1 summarizes the components of the Twitter News Model (TNM) with its news beliefs, motivations and attitudes as they are characterized in the context of Twitter, along with a Twitter-specific definition of news consumption. In the next section, we describe how the TNM can be leveraged to predict and explain news consumption.

5 REALIZING THE TWITTER NEWS MODEL

In this section, we apply the proposed Twitter News Model (TNM) to a dataset of interactions between users and journalists/newspapers to elucidate the dynamics of news consumption on Twitter. Specifically, this empirical study is designed to (i) reveal *what* drives users' consumption of news on Twitter, and (ii) predictively relate users' news beliefs, motivations, and attitudes to their consumption of news on Twitter. We characterize the various components of the model – news beliefs, motivations and attitudes – and relate them to the computation of consumption.

Dataset. We collected all tweets posted by 200 manually curated Irish media sources and journalists accounts covering 79 different news outlets.[2] The accounts were selected to cover all major national and regional media outlets for Ireland, in addition to individual journalists writing for these outlets. From these 200 accounts, 117 are individual journalists' (31 female and 86 male) and 83 are corporate accounts. In addition, we collected all the retweets and mentions received by these 200 Twitter accounts. Previous work has shown that the use of Twitter for news in Ireland is broadly similar to that found in other western countries (e.g., the UK and France [28]). The data collection spans a period of four months from August 10th to December 10th, 2017. This period covers a series of news events including Hurricane Harvey, the North Korea's launch of missiles, and the Las Vegas shooting. In total, 16.6K users retweeted and/or mentioned one or more of the 200 journalists and newspapers under study, producing a total of 352,500 (μ = 22, σ = 48) interactions.

# News media accounts	200 (117 individual and 83 corporate)
# Interactions	352,500 (tweets with mentions and retweets)
# Users	16,683 unique users
Time period	from August 10th to December 10th, 2017

Table 2: Dataset statistics.

As our aim is to understand what makes audiences interact with news-tweets, journalists, and newspapers in Twitter, we focus on the 352.5K interactions (retweets and tweets with mentions) posted by the 16.6K users. We do not analyze the tweets posted, nor the interactions started by our 200 journalistic accounts.[3] Table 2 summarizes the main statistics of our Twitter corpus.

News Beliefs. News beliefs are defined in terms of user mentions of journalists and newspapers, more than merely retweeting their news-tweets (see Equation 1). In our corpus, we found a total of 52% (8.7K) of users who post at least as many tweets with mentions as they do retweets (i.e., $B(u) = 1$). Of these, 4.3% (713) of users post retweets and mentions in a 50/50 ratio.

Users mention journalists or newspapers for different reasons, including discussing topics of interest (e.g., *@IrishTimes Each year the housing prices rise. How do we rebuild at a rate of such increase? We can't. Stop housing...*), expressing opinions and concerns (e.g., *@gavreilly 2017 results at home terrible. Two draws and defeat from Wales, Austria and Serbia cost great chance to top the group*), or simply to establish conversations (e.g., *@MiguelDelaney Really great piece Miguel, fair play. I think the FAI need to import some bright minds from Germany...*). Users who mention news providers more often, spend time and put effort into tailoring their tweets, which could indicate that news (and its discussion), is valuable to them. The remaining 48% (7945) of users in our corpus, retweet more than they mention (i.e., $B(u) = 0$); and 54% (4330) of them do so with a retweets-mentions ratio of at least 80/20.

These findings indicate that for the majority of users in our dataset, news is important to the extent in which they feel the urge to not only spread tweets from journalistic sources, but to express their own opinions, post their own tweets, and become active participants in the discussion of news.

News Motivations. News motivations are what drive people to consume news. In order to better understand the reasons behind the news consumption for each user in our corpus, we separated her/his interactions (i.e., retweets + mentions) by news category.

Our model proposes that the proportion of interactions a user has with a given news category, indicates their motivation for consuming news. For example, if a user only retweets sports news-tweets and/or mentions sports journalists, then this indicates they are largely consuming news for entertainment purposes, in contrast to the motivations of a user that only interacts with news-tweets/journalists in the business category.

To identify the news category of the tweets posted by the users in our corpus, we followed the process described in [29]. We first

[2]We use Twitter streaming API. Every 30 minutes, the crawler collected the new tweets posted and received by these accounts of interest.

[3]Note, that our data collection only includes the interactions between users and journalistic accounts, during the time-period under study; these users' interactions with other users (i.e., non-journalists), were not collected.

News Category	Accounts
Business	13 (11%)
Lifestyle	15 (13%)
Breaking News	30 (26%)
Politics	25 (21%)
Science and Technology	6 (5%)
Sports	28 (24%)
Total	117

Table 3: News categories and corresponding number of individual journalists' accounts.

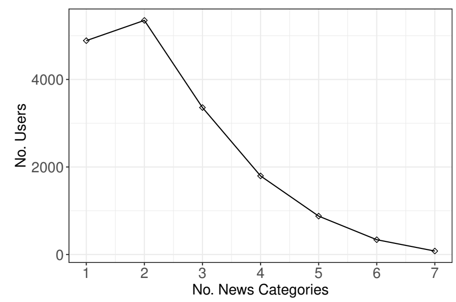

Figure 2: Audience vs. number of different news categories they interact with.

separated the 200 Twitter accounts into individual and corporate. Corporate accounts (e.g., *@Independent_ie*) post tweets that span all news categories, while individual accounts (e.g., *@conor_pope*) tend to focus on a specific area, such as sports or business. Three annotators judged the news category of each individual account based on (i) a random sample of 50 tweets sent by the individual journalist, and (ii) a list of the top-100 terms used by the journalist in her/his tweets during the period of interest. The news categories considered are business, lifestyle, science and technology, breaking news, politics, and sports.[4] We follow this human annotation approach, so we could obtain high quality labels that can be used as ground truth in related future tasks. The distribution of accounts across the six news categories is shown in Table 3. We group the remaining 83 Twitter accounts under a category that we will call *corporate*, as they cannot be classified under any specific news category. The Fleiss' Kappa inter-rater agreement over all news categories is $\kappa = 0.51$.

The interactions between a user and a news category were calculated as follows: if user u mentions a journalist or retweets a news-tweet from a journalist who belongs to news category c (e.g., sports), then we add 1 to user u's interactions with c. We follow the same procedure for all user u's tweets. We then normalize the total number of user u's interactions per category (i_{uc}) by the total number of user u's interactions over all news categories (I_u). We repeat this procedure for all users (U) in our corpus.

Using this analysis a number of interesting regularities can be found in the way users interact with news categories. Figure 2 shows the number of users in our corpus who interact with one or more news categories. Out of the 16.6K users, 4887 (29.3%) interact (i.e., retweet and mention journalists) solely with one news category. The three most popular categories among these users are corporate (3383 users only interact with corporate accounts), sports (1134 users), and politics (220 users). A further third of this audience (5350 or 32%) interacts exclusively with two news categories. The most popular pairs of news categories are corporate-sports (1882 users only interact with corporate and sports accounts, with 717 users interacting more with sports than with corporate accounts, and 1165 doing the opposite), corporate-politics (1847 users), and corporate-breaking news (843).

Of the remainder, 3358 (20.1%) of users interact with three different news categories. The most popular triplet for these users is politics-corporate-breaking news (1074 users only interact with these three news categories). From four to seven news categories, we observe a sharp decrease in the number of users. Only, 1796 (10.8%) users interact with four news categories, 877 (5.3%) with five, 338 (2%) with six, and 79 (0.5%) with all the seven news categories.

These findings show that, for this user cohort, the majority of people who follow news on Twitter tend to interact with 1-3 news categories. Corporate accounts (e.g., *@Independent_ie*) seem to be a major focus of user attention and interactions, independently of the other specific news categories that a user follows. One possible reason for this behavior, may be that audiences in Twitter are consuming news as a source of general knowledge, and/or to keep up with events of public interest. Sport and politics are the two other news categories that receive considerable attention. This result evidences more niche interests, in which users have a news category of choice, that presumably fulfills more fine-grained motives such as entertainment (e.g., sports news), or keeping an up-to-date knowledge on government-related affairs (e.g., political news).

News Attitudes. News attitudes reflect users overall orientation to the consumption of news and, in our model is measured by extracting the polarity of the users' mentions (i.e., tweets in which users mention journalists or newspapers). This means that for a given user, we first need to determine the polarity of her individual tweets and then average the polarities of all her tweets. These aims were achieved in three steps: (i) all the tweets of concern were transformed into vector representations using a word-embedding technique, (ii) a polarity classifier was trained on a corpus of tweets with known polarities, (iii) this classifier was used to find the polarity score for each tweet of a given user and then these scores were averaged over all the user's tweets. Each of these three steps is detailed further as follows.

(i) Transforming tweets into word vectors. Several word-embedding techniques have been proposed to transform words into vector representations (e.g., *Word2Vec* [25], and *GloVe* [30]). The tweets of concern here were processed using GloVe, an unsupervised learning algorithm that provides a collection of word vectors pre-trained on 2B tweets/27B tokens [30]. These vectors are available in different dimensionalities: 25, 50, 100, and 200 dimensions.[5] Using the 200-dimension vectors, we found the GloVe for each word in a

[4]Note that the same news categories can have different names depending on the news provider, we show here particularly representative ones.

[5]https://nlp.stanford.edu/projects/glove/

tweet,[6] and averaged all these word vectors to obtain the overall tweet vector.

(ii) Training a polarity classifier. To train a tweet polarity classifier[7], we used the word vectors (i.e. GloVe) representation of the SemEval-2017[8] dataset [33] consisting of 19.9K tweets. These tweets were posted between 2013 and 2016 and had identified polarities: 8157 positive, 8101 negative, and 3079 neutral tweets. In training the classifier, only positive and negative tweets were used; we used upsampling on the negative tweets to ensure a balanced dataset (i.e., 8157 positive and 8157 negative tweets). The classifier was trained to predict the probability that a given tweet belonged to the positive class (i.e., the closer to 1 the more probable the tweet was positive, the closer to 0, otherwise). Using 5-fold cross validation, the final classifier had an average F1 measure of $F1 = 0.78$. The SemEval datasets are widely used for sentiment analysis tasks [7, 20, 34]. An interesting extension to polarity classification, would be to analyze the presence of other emotions known to be expressed in tweets, such as anger, joy, surprise, or sadness, as done in [26, 27].

(iii) Aggregating the polarity of a user's tweets. For each user u in our corpus, we extract the polarity for each individual tweet in which user u mentioned a journalist or a newspaper. As a final step, we average the polarity scores of all user u's tweets (i.e., tweets with mentions) and assign that average as the overall user's polarity.

Out of the 16.6K users in our corpus, 13,097 (78.5%) show a positive attitude ($A(u) \geq 0.5$) toward journalists and newspapers; of these users, 190 had a strongly positive polarity average ($A(u) > 0.8$). The remaining 3,578 (21.5%) users show an average attitude that is more inclined toward a negative polarity ($A(u) < 0.5$); of these users, 62 had a strongly negative polarity average ($A(u) < 0.2$). Although we find some news consumers to be more negative than positive, the majority of people show a positive attitude toward news and news providers.

News Consumption. News consumption is captured by the total number of interactions (i.e., retweets + mentions) that user u has with news-tweets, journalists, and newspapers. Here, we found that on average, each of our 16.6K users interacts 21.3 times ($\sigma = 48$) with news-tweets/news providers. Of these users, 508 (3%) interact 100 times or more, 9,303 (56%) users interact 10 times or less, and the remaining 6,874 (41%) of users interact between 11 and 99 times in the four-month period spanned by our data collection. These results show that 41% of the Twitter users in our dataset, turn to news on a regular basis. Some of these users tend to do so approximately once a day. A small group (3%), interacts with news tweets and/or news providers on average 4 times a day, with some users retweeting and mentioning as much as 13 times a day, on average. For more than half of our news consumers (56%), interacting with news-tweets and news providers seems to be more of an sporadic activity.

5.1 Predict & Explain News Consumption

Having captured the behavior of our news consumers using the Twitter News Model, we turn our attention to building a predictive model to better understand the drivers of news consumption. We cast the news consumption prediction as a regression task in which our goal is to predict, based on users' features, their corresponding news consumption. Our aim is to train a machine learning regressor to predict the news consumption, which in our model corresponds to the *number of interactions (i.e., retweets + mentions)* between each user and news-tweets/news providers (cf. Section 4). We use the z-score of the users' interactions as our target variable.

User representation. Each user in our data collection is characterized using a feature vector. Users' features are extracted using three methods described as follows:

- *Twitter News Model (TNM).* We construct users' vectors using the TNM mediators; these are 9-dimensional vectors with one dimension for news beliefs (i.e., whether the user tends to mention more than retweet), seven for news motivations (i.e., one per each news category and one for corporate accounts), and one for news attitudes (i.e., the polarity score for the user).

- *Word-Embeddings (WE).* Using the same procedure described in Section 5, we construct feature vectors based on the content of the users' interactions. For each user, we averaged the GloVe word-embeddings [30] of her/his retweets and mentions to obtain a 200-dimensional vector that is fully based on the content of the user's tweets.[9]

- *Twitter News Model+Word-Embeddings (TNM+WE).* Vectors combining the Twitter News Model and the word-embeddings representations of the tweets. This results in vectors of 209 dimensions: 9 from the TNM and 200 from the word-embedding representations.

Experimental setup. We split our data into training and test sets using an 80%/20% ratio. We explore three different regression models, namely Random Forest (RF) [11], Gradient Boosting Trees (GBT) [11], and Extremely Randomized Trees (ET) [10] and grid-search their best hyperparameters using 10% of the training set for validation. We conduct 25 rounds of experiments and measure the prediction quality of our models using the *Mean Squared Error (MSE)* metric. The results reported are the average MSE on the test set over the 25 rounds. Based on this exploration, the model that exhibits the best predictive performance for all three user representations is Extremely Randomized Trees (ET),[10] and therefore, the one chosen for our task. ET fits a number of randomized decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The best hyperparameters found for our task are 1000 estimators (n_estimators=1000) and a minimum of 3 samples required to be at a leaf node (min_samples_leaf=3).

We train three separate ET regression models, one per each set of features extracted, namely TNM, WE, and TNM+WE. Figure 3 shows each model's average MSE, whose values correspond to 0.587 for TNM, 0.439 for WE, and 0.433 for TNM+WE.

Interpretability vs. predictive power. As we observe from Figure 3, the model trained using the TNM features has the highest error. In terms of predictive power, the Word-embeddings and TNM+Word-embeddings models are better. However, the model based on the Twitter News Model has one overriding benefit; namely, that it is fully interpretable, in that we know the semantics of the

[6] If the word does not have a GloVe representation, we set to 0 all the 200 dimensions of the corresponding vector
[7] scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[8] SemEval (Semantic Evaluation) is an ongoing series of evaluations of computational semantic analysis systems, organized under the umbrella of SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics.
[9] Unlike in the calculation of News Attitudes (Section 5), these vectors also take into account users' retweets, not just their original tweets with mentions.
[10] scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html

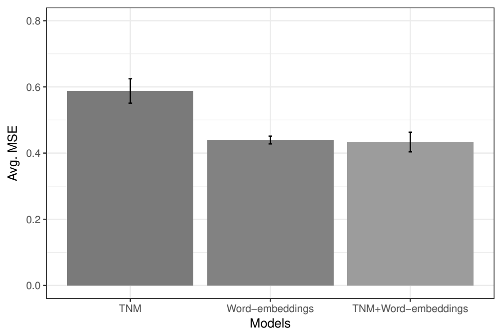

Figure 3: Average MSE values for the different models in predicting news consumption, showing 95% confidence intervals (as these are error values, the lower the better).

corresponding TNM feature. Notably, the hybrid TNM+WE model, combines WE's predictive power with TNM's interpretability, allowing us to produce predictions of comparable quality (to those produced by the WE model alone), while identifying the meaningful features that explain these predictions. We cannot obtain the same interpretability from the WE model on its own given its nature, i.e., lack of semantics of the word embeddings.

From our hybrid model (TNM+WE), we extracted the features that were most important (to the ET regressors) in the prediction of news consumption. Interestingly, the top-2 features are TNM's news motivations, specifically, users' interaction with the news categories business and breaking news. In rank 3, we find a content feature (i.e., from the word-embeddings representation), followed by two other features of news motivations in ranks 4 and 5, the interaction with corporate accounts and the politics news category. The remaining features in the top-20 are content features, with the exception of the interaction with lifestyle news in position 13. For our hybrid model, the TNM's features considered less important for prediction were news attitudes (i.e., the polarity of users' interactions), users' interactions with the news categories of science and technology and sports (remaining features of news motivations), and news beliefs (i.e., whether the users tend to mention more than to retweet).

5.2 Discussion

By applying the Twitter News Model to our dataset, we were able to extract interesting insights into Twitter users' news consumption. We found that the most telling features in the prediction of users' news consumption are *news motivations*, or the reasons that drive them to consume news. This finding is consistent with the findings of the Motivational Consumption Model, in which the authors explain that news motivations influence both news attitudes and news consumption. While we are now able to indicate that users' interactions with certain news categories (e.g., business and politics) are useful in predicting their news consumption, further analyses are necessary in order to explain whether this means they consume more or less news than if they interacted with other news categories.

Although most of the observed effects can be attributed to news motivations, *news attitudes* also play a role. We found that the cohort in our sample tends to approach journalists with a positive attitude, as reflected in the polarity of the tweets in which they mention journalistic accounts.

The features reflecting *news beliefs* appeared to have the least importance in the prediction of news consumption. This finding is in line with the MCM, in which *news beliefs* are seen as having a direct effect on news motivations and news attitudes, and only through these, showing an effect on news consumption itself. This result might suggest that users who are active participants in the news conversation in Twitter, implicitly convey the value that news has for them, by virtue of such interactions (which explicitly reflect their news motivations and attitudes).

Using the TNM features alone to learn and predict users' news consumption, did not result in the best model; as the predictions had the highest error. Other approaches, including those using content-based features, for example word-embeddings, have higher predictive power. However, the TNM models help make results more interpretable. This characteristic is essential, particularly in environments where beyond predicting news consumption, is the knowledge obtained in the process that matters, for example, if journalists learn what makes Twitter users consume news, they can apply the knowledge to future interactions and thus, keep their audience and/or possibly attract more readers. For tasks such as news recommendation and personalization, journalists can use these results to segment their audience by creating *personas* based on the news consumption behavior. News can then be tailored and targeted to each persona, highlighting the aspects that will make it more likely to be consumed.

6 CONCLUSION

We have presented our Twitter News Model (TNM), a computational approach to modeling the news consumption behavior of Twitter users. Our approach is based on the Motivational Consumption Model (MCM), developed to model the consumption of news in conventional media such as newspapers, news aggregators, and TV. To the best of our knowledge, this is the first time that MCM has been adapted to this social media context. We started by discussing the theoretical background on which the MCM was based, described how we used the theory to instantiate our Twitter-specific model for news consumption (TNM), showed the usefulness of our model by applying it to a real-world scenario in which we learn the dynamics of news consumption in Twitter, and used this knowledge in a news consumption prediction task. Our findings reveal that news motivations, followed by news attitudes, and news beliefs, impact users' behavior of news consumption in Twitter. In future work, we aim to apply our findings to tasks such as user segmentation, and to inform strategies that journalists might use to promote their news on Twitter, so that they can better understand the dynamics of news consumption on social media. In addition, we aim to explore other interpretable user representations that can be used in prediction tasks related to the one addressed in this paper.

ACKNOWLEDGMENTS

The authors would like to thank *The Irish Times* for their funding and help on this project. This work is supported by Science Foundation Ireland through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289.

REFERENCES

[1] 2015. *Reuters Institute Digital News Report*. Technical Report. http://www.digitalnewsreport.org/

[2] Pablo Boczkowski, Eugenia Mitchelstein, and Mora Matassi. 2017. Incidental News: How Young People Consume News on Social Media.. In *HICSS*. AIS Electronic Library (AISeL).

[3] Alex Bruns, Brenda Moon Moon, Felix Münch, Jan-Hinrik Schmidt, Lisa Merten, Hallvard Moe, and Sander Schwartz. 2016. NEWS SHARING ON TWITTER: A NATIONALLY COMPARATIVE STUDY. *Selected Papers of AoIR 2016: The 17th Annual Conference of the Association of Internet Researchers* 0, 0 (2016).

[4] Pew Research Center. 2014. How social media is reshaping news. http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/

[5] Boreum Choi and Inseong Lee. 2017. Trust in open versus closed social media: The relative influence of user- and marketer-generated content in social network services on customer trust. *Telematics and Informatics* 34, 5 (2017), 550–559.

[6] Sophie Chou and Deb Roy. 2017. Nasty, Brutish, and Short: What Makes Election News Popular on Twitter?. In *ICWSM*. AAAI Press, 492–495.

[7] Mathieu Cliche. 2017. BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs. *arXiv preprint arXiv:1704.06125* (2017).

[8] Stephanie Edgerly, Kjerstin Thorson, Esther Thorson, Emily K Vraga, and Leticia Bode. 0. Do parents still model news consumption? Socializing news use among adolescents in a multi-device world. *New Media & Society* 0, 0 (0), 1461444816688451. https://doi.org/10.1177/1461444816688451 arXiv:https://doi.org/10.1177/1461444816688451

[9] Martin Fishbein and Icek Ajzen. 2010. *Predicting and Changing Behavior*. Psychology Press.

[10] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely Randomized Trees. *Mach. Learn.* 63, 1 (April 2006), 3–42. https://doi.org/10.1007/s10994-006-6226-1

[11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*. Springer New York Inc., New York, NY, USA.

[12] Benjamin D. Horne and Sibel Adali. 2017. The Impact of Crowds on News Engagement: A Reddit Case Study. *CoRR* abs/1703.10570 (2017). arXiv:1703.10570 http://arxiv.org/abs/1703.10570

[13] JAMES JACCARD. 2012. The Reasoned Action Model: Directions for Future Research. *The Annals of the American Academy of Political and Social Science* (2012), 58–80.

[14] Bente Kalsnes and Anders Olof Larsson. 2017. Understanding News Sharing Across Social Media. *Journalism Studies* 0, 0 (2017), 1–20. https://doi.org/10.1080/1461670X.2017.1297686 arXiv:https://doi.org/10.1080/1461670X.2017.1297686

[15] Yaser Keneshloo, Shuguang Wang, Eui-Hong Sam Han, and Naren Ramakrishnan. 2016. Predicting the Popularity of News Articles.. In *SDM*, Sanjay Chawla Venkatasubramanian and Wagner Meira Jr. (Eds.). SIAM, 441–449.

[16] Yaser Keneshloo, Shuguang Wang, Eui-Hong Sam Han, and Naren Ramakrishnan. 2016. Predicting the shape and peak time of news article views.. In *BigData*, James Joshi, George Karypis, Ling Liu, Xiaohua Hu, Ronay Ak, Yinglong Xia, Weijia Xu, Aki-Hiro Sato, Sudarsan Rachuri, Lyle H. Ungar, Philip S. Yu, Rama Govindaraju, and Toyotaro Suzumura (Eds.). IEEE, 2400–2409.

[17] Yonghwan Kim, Youngju Kim, Yuan Wang, and Na Yeon Lee. 2016. Uses and Gratifications, Journalists' Twitter Use, and Relational Satisfaction with the Public. *Journal of Broadcasting & Electronic Media* 60, 3 (2016), 503–526. https://doi.org/10.1080/08838151.2016.1164171 arXiv:https://doi.org/10.1080/08838151.2016.1164171

[18] Huyen T. Le, Zubair Shafiq, and Padmini Srinivasan. 2017. Scalable News Slant Measurement Using Twitter.. In *ICWSM*. AAAI Press, 584–587.

[19] Angela M Lee and Hsiang Iris Chyi. 2014. Motivational consumption model: Exploring the psychological structure of news use. *Journalism & Mass Communication Quarterly* 91, 4 (2014), 706–724.

[20] Quanzhi Li, Sameena Shah, Armineh Nourbakhsh, Rui Fang, and Xiaomo Liu. 2017. funSentiment at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs Using Word Vectors Built from StockTwits and Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 852–856.

[21] Maxwell McCombs and Paula Poindexter. 1983. The Duty to Keep Informed: News Exposure and Civic Obligation. *Journal of Communication* 33, 2 (1983), 88–96. https://doi.org/10.1111/j.1460-2466.1983.tb02391.x

[22] Pasquale De Meo, Fabrizio Messina, Domenico Rosaci, and Giuseppe M. L. SarnÃĪ. 2017. Forming time-stable homogeneous groups into Online Social Networks. *Inf. Sci.* 414 (2017), 117–132.

[23] Pasquale De Meo, Katarzyna Musial-Gabrys, Domenico Rosaci, Giuseppe M. L. SarnÃĺ, and Lora Aroyo. 2017. Using Centrality Measures to Predict Helpfulness-Based Reputation in Trust Networks. *ACM Trans. Internet Techn.* 17, 1 (2017), 8:1–8:20.

[24] Siakalli Michailina, Andreas Masouras, and Christos Papademetriou. 2015. Uses and Gratifications in Online News: Comparing Social Media and News Media Use by Users. (05 2015).

[25] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality.. In *NIPS*, Christopher J. C. Burges, LÃĪon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger (Eds.). 3111–3119.

[26] Claudia Orellana-Rodriguez, Ernesto Diaz-Aviles, and Wolfgang Nejdl. 2013. Mining emotions in short films: user comments or crowdsourcing?. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 69–70.

[27] Claudia Orellana-Rodriguez, Ernesto Diaz-Aviles, and Wolfgang Nejdl. 2015. Mining affective context in short films for emotion-aware recommendation. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 185–194.

[28] Claudia Orellana-Rodriguez, Derek Greene, and Mark T. Keane. 2016. Spreading the News: How Can Journalists Gain More Engagement for Their Tweets?. In *Proceedings of the 8th ACM Conference on Web Science (WebSci '16)*. ACM, New York, NY, USA, 107–116. http://doi.acm.org/10.1145/2908131.2908154

[29] Claudia Orellana-Rodriguez, Derek Greene, and Mark T. Keane. 2017. Spreading One's Tweets: How Can Journalists Gain Attention for their Tweeted News? *J. Web Science* 3, 2 (2017), 16–31.

[30] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.

[31] Alicja Piotrkowicz, Vania Dimitrova, Jahna Otterbacher, and Katja Markert. 2017. Headlines Matter: Using Headlines to Predict the Popularity of News Articles on Twitter and Facebook.. In *ICWSM*. AAAI Press, 656–659.

[32] Julio C. S. Reis, Haewoon Kwak, Jisun An, Johnnatan Messias, and FabrÃŋcio Benevenuto. 2017. Demographics of News Sharing in the U.S. Twittersphere.. In *HT*. ACM, 195–204.

[33] Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval '17)*. Association for Computational Linguistics, Vancouver, Canada.

[34] Alon Rozental and Daniel Fleischer. 2017. Amobee at SemEval-2017 Task 4: Deep Learning System for Sentiment Detection on Twitter. *CoRR* abs/1705.01306 (2017). arXiv:1705.01306 http://arxiv.org/abs/1705.01306

[35] Terrence Szymanski, Claudia Orellana-Rodriguez, and Mark T Keane. 2017. Helping News Editors Write Better Headlines: A Recommender to Improve the Keyword Contents & Shareability of News Headlines. *arXiv preprint arXiv:1705.09656* (2017).

[36] Beidou Wang, Can Wang, Jiajun Bu, Chun Chen, Wei Vivian Zhang, Deng Cai, and Xiaofei He. 2013. Whom to Mention: Expand the Diffusion of Tweets by @ Recommendation on Micro-blogging Systems. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 1331–1340. https://doi.org/10.1145/2488388.2488505

[37] Bo Wu and Haiying Shen. 2015. Analyzing and predicting news popularity on Twitter. *Int J. Information Management* 35, 6 (2015), 702–711.