

# Tabloids in the Era of Social Media? Understanding the Production and Consumption of Clickbaits in Twitter

ABHIJNAN CHAKRABORTY, Indian Institute of Technology Kharagpur, India  
RAJDEEP SARKAR, Indian Institute of Technology Kharagpur, India  
AYUSHI MRIGEN, Indian Institute of Technology Kharagpur, India  
NILOY GANGULY, Indian Institute of Technology Kharagpur, India

---

With the growing shift towards news consumption primarily through social media sites like Twitter, most of the traditional as well as new-age media houses are promoting their news stories by tweeting about them. The competition for user attention in such mediums has led many media houses to use catchy sensational form of tweets to attract more users – a process known as *clickbaiting*. In this work, using an extensive dataset collected from Twitter, we analyze the social sharing patterns of clickbait and non-clickbait tweets to determine the organic reach of such tweets. We also attempt to study the sections of Twitter users who actively engage themselves in following clickbait and non-clickbait tweets. Comparing the advent of clickbaits with the rise of *tabloidization* of news, we bring out several important insights regarding the news consumers as well as the media organizations promoting news stories on Twitter.

CCS Concepts: • **Human-centered computing** → **Social content sharing; Social media; Empirical studies in collaborative and social computing;**

Additional Key Words and Phrases: Clickbaits; Online News; News Sharing in Twitter; Clickbait Readership; Tabloids; Public Sphere; Twitter User Demographics

## ACM Reference Format:

Abhijnan Chakraborty, Rajdeep Sarkar, Ayushi Mrigen, and Niloy Ganguly. 2017. Tabloids in the Era of Social Media? Understanding the Production and Consumption of Clickbaits in Twitter. Proc. ACM Hum.-Comput. Interact. 1, CSCW, Article 30 (November 2017), 21 pages. <https://doi.org/10.1145/3134665>

---

## 1 INTRODUCTION

Historically, news has been a very important part of societal evolution, playing an integral role in the same since the 17th century [52]. In the years that followed, the medium has only increased the power it wields over the perceptions and beliefs of the general public. The press has been considered as one of the fundamental pillars of any functioning democracy for a considerable period of time, and the editors of news organizations have long viewed their roles as the custodians of society, deciding which news should be consumed by the common people, and which shouldn't [49].

---

We thank the anonymous reviewers whose suggestions helped to improve the paper. We also thank Bhargavi Paranjape and Adarsh Pandey for being with us in the initial phase of this work. Abhijnan Chakraborty is a recipient of Google India PhD Fellowship and Prime Minister's Fellowship Scheme for Doctoral Research, a public-private partnership between Science & Engineering Research Board (SERB), Department of Science & Technology, Government of India and Confederation of Indian Industry (CII).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

2573-0142/2017/11-ART30

<https://doi.org/10.1145/3134665>

The first disruption in the media landscape came in the late 20th century from the *tabloidization of news*, where detailed truthful reporting of important events gave way to stories that were more sensational in nature. Since its inception, the tabloid style has been criticized as inferior, appealing to base instincts and public demand for sensationalism [6].

In the 21st century, online news consumption has gained momentum, and the revenue model has shifted towards being mostly advertisement driven, where users do not have to pay anything to read the news articles and the money comes from the clicks on the advertisements that are present on the news websites. Such shift in news consumption has caused a significant change in both the consumption pattern as well as the means of offering news content to the readers.

The attractiveness of a news article's content has become more relevant than the credibility of the organization writing that article, and news readers now have a negligible cost for switching from one media source to another. This, in turn, has resulted in a fierce competition between the news outlets to capture the readers' attention. A by-product of this competition is the advent of **clickbaits**, where in order to tempt the readers to click and read articles in their websites, the media organizations use flashy headlines that pique the interest of the readers [14].

Clickbaits are typically defined as the headlines that are intended to lure readers, by providing a small glimpse of what to expect from the article [9]. According to the Oxford English Dictionary, clickbait is "*(On the Internet) content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page*"<sup>1</sup>. Examples of clickbaits include "17 Reasons You Should Never Ever Use Makeup", "These Dads Quite Frankly Just Don't Care What You Think", or "10 reasons why Christopher Hayden was the worst 'Gilmore Girls' character".

On one hand, the success of such clickbait headlines in attracting visitors to the news websites has propelled new age digital media companies like BuzzFeed to have valuation three times that of century old The Washington Post [36]. However, on the other hand, the articles themselves often offer less news value, and therefore, concerns have been raised by traditional media organizations regarding the role of *journalistic gatekeeping* in the era of clickbaits [19, 22].

The rise of clickbaits in online news bears many similarities with the advent of tabloids. In general, there has been a lot of discussion about the positive and negative effects of tabloidization in the traditional media. Many researchers believe that tabloids have had a significant role to play in lowering the standards of news. According to Colin Sparks, "*Public ignorance and apathy is growing as the serious, challenging and truthful is being pushed aside by the trivial, sensational, vulgar and manipulated*" [57]. However, on the other hand, other researchers have argued that *softening* of news by the tabloids helped raising political awareness among politically inattentive citizens [5, 17]. There have been long drawn debates around the content of tabloids, and the effects of the same on the *public sphere* [38]. In spite of all the hue and cry against tabloidization, it has been agreed that blindly disregarding the importance of the tabloid form of journalism is not advisable [38].

Contrary to the study of tabloids, research on clickbaits is still in a very nascent stage, where recently a few attempts have been made to automatically detect clickbait headlines in different media sites [3, 14, 26, 40]. However, to our knowledge, there has been no attempt to understand the consumers who follow such headlines, and further help in expanding the reach of the corresponding articles to a wider audience. In fact, all past works mostly highlight the negative aspects of clickbaits without considering their consumers [14, 15].

Considering the consumers of news is especially important today because there has been another parallel shift in news consumption over the last few years, where social media has become the primary medium for news consumption [34]. This has led to two other disruptions for the media organizations.

---

<sup>1</sup>[oxforddictionaries.com/us/definition/american\\_english/clickbait](http://oxforddictionaries.com/us/definition/american_english/clickbait)

First, in social media, anyone can become a publisher of content with virtually zero upfront cost, which has led to mushrooming of several *social media only* publisher startups [4]. Emergence of such organizations has further crowded the media landscape already flooded with national, international and local news outlets, resulting in immense competition for user attention in such mediums.

Second, the media organizations also need to adapt to another news dissemination medium, specifically related to the posting rules of that particular medium. A prime example of this is Twitter, which severely limits the number of characters that a single tweet can contain, and thus imposing an upper bound on the number of words that the news outlets can use to attract the users.

Such drastic changes in the medium of propagation of news, and the competition for user attention in such mediums have led many media houses to use catchy sensational posts to attract more users. We broaden the original definition of clickbaits to include such catchy social media posts, which not only encourage users to click on the embedded news article links, but also persuade users to share these posts with their peers, which in turn help in increasing the media house's follower base. With the growing adoption of clickbaiting techniques, and social news consumption leading to transformational changes in the media landscape, we believe that at this point, a detailed study of clickbaits and their consumers in social media is required to get the holistic picture.

In this paper, we take a step towards that direction by analyzing the users as well as the usage of clickbaits in social media sites. We collect extensive longitudinal data over eight months from the popular social media Twitter covering both clickbait and non-clickbait (or traditional) tweets, and then attempt to explore several important dimensions related to these two types of tweets and their users.

More specifically, in this paper, we attempt to investigate the following research questions:

**RQ1. How are clickbait tweets different from non-clickbait tweets?**

**RQ2. How do clickbait production and consumption differ from non-clickbaits?**

**RQ3. Who are the consumers of clickbait and non-clickbait tweets?**

**RQ4. How do the clickbait and non-clickbait consumers differ as a group?**

Our investigation reveals several interesting insights on the production of clickbaits. For example, clickbait tweets include more entities such as images, hashtags, and user mentions, which help in capturing the attention of the consumers. Additionally, we find that a higher percentage of clickbait tweets convey positive sentiments as compared to non-clickbait tweets. As a result, clickbait tweets tend to have a wider and deeper reach in its consumer base than non-clickbait tweets.

We also make multiple interesting observations regarding the consumers of clickbaits. For example, clickbait tweets are consumed more by women than men, as well as by more younger people compared to the consumers of non-clickbaits. Additionally, they have higher mutual engagement among each other. On the other hand, non-clickbait consumers are more reputed in the community, and have relatively higher follower base than clickbait consumers.

Comparing clickbaits with tabloids, we find commonalities (e.g., both make optimal use of graphic elements) as well as differences (e.g., clickbaits tend to convey more positive sentiments vs more negativity in tabloid reporting). Likewise, clickbait and tabloid readership have similarities (e.g., both cater to younger population) and differences (e.g., majority of tabloid readers are male, which is in contrast to the female dominance among clickbait consumers).

In summary, we make two major contributions in this paper: (i) to our knowledge, ours is the first attempt to understand the consumers of clickbaits, and (ii) while doing so, we also make the first effort to contextualize the rise of clickbaits with the tabloidization of news. As we mentioned earlier, all past works on clickbaits only highlight its negative aspects. We believe that our work can foster further research, raise debates in the community, and help bring in a more holistic view of the entire spectrum.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Tabloidization and its impact on news media

Clickbaits can be thought as the digital successor to the *tabloidization* of print journalism [50]. Tabloids disrupted a long-held approach towards *journalistic gatekeeping* by focusing more on *soft news* than *hard news*, and on sensationalizing the content over the detailed truthful reporting of events. There have been concerns in the journalism community regarding the tabloidization of news and its potential threat to democracy [45, 50]. However, on the other hand, several studies have noted that *softening* of news by the tabloids helped raising political awareness among politically inattentive citizens [5, 17].

Likewise, with the sudden increase in the prevalence of clickbaits in the digital media landscape, similar concerns have been raised. There have been different discussions constantly reprimanding the low news value of clickbaits, and the change in the face of journalistic gatekeeping that clickbaits bring with them [19]. In this paper, we try to tie the existing works on tabloids, and see whether some of the arguments made there still hold for clickbaits. We also attempt to argue for a holistic debate about the usefulness of clickbaits, similar to the one present for tabloids.

### 2.2 Readership of traditional and tabloid newspapers

Traditionally, to decide the correct audience for advertising goods and services, advertisers as well as market research agencies conducted readership surveys for offline newspapers. In such surveys, it has been found that the traditional broadsheet newspapers mostly cater to the affluent and well educated audience, with majority of readers in the upper middle class and middle class professional and managerial social grades (denoted as grade AB by National Readership Survey<sup>2</sup> in the UK) [29]. Whereas, the majority of the tabloid readers are found to be in the C1C2-E social grades, i.e., people in lower middle and working class, involved in supervisory, clerical, or skilled and unskilled manual works [43]. Thus, there seems to be a clear distinction between the consumers of tabloids and traditional news. Following this line of work, in this paper, we investigate whether the clickbait and non-clickbait consumers in Twitter also differ substantially.

### 2.3 Tabloids and the public sphere

Most of the criticisms around tabloidization are grounded in the notion of the *public sphere*, developed in the seminal work by Habermas [27]. According to Habermas, public sphere is a realm of our social life where public opinion can be formed via *rational-critical debate between private individuals on public matters* [27], and news media is the primary enabler of such communications. By such construction, Habermas puts media in the normative center of a well-functioning democracy, and hence the standard of the content propagated by the media becomes immensely important [20]. Critiques of tabloids have argued that tabloids fail those standards to enable debates in the public sphere [29]. In this work, we investigate a related question in the context of social media – among clickbaits or non-clickbaits, which is a better enabler of communication between different groups of people in the social media, which is the public sphere in this context.

### 2.4 Social media and journalistic gatekeeping

The wide-spread adoption of social media sites like Facebook and Twitter has led to a paradigm shift in how news stories are consumed by people world-wide. Today, a large and growing fraction of news readers are finding news stories on these social media platforms. Unlike the traditional media landscape, the newspaper editors no longer exert editorial control over the type of stories getting shared, and hence, there is little journalistic gatekeeping applied on the news discourse.

<sup>2</sup>[nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade](https://nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade)

A lot of prior works have discussed the effect of such drastic change in the news landscape. By studying the sharing patterns of 8,000 news items across major news outlets, Diakopoulos *et al.* [18] concluded that the newsreaders in social media act as *network gatekeepers*, and more often (re)share news items on socially deviant events. Chakraborty *et al.* [12] analyzed how such network gatekeeping role differs across different sharing mediums. In another work, they observed that the gatekeeping roles exercised by both traditional and network gatekeepers can lead to temporal coverage bias [11]. Further, Matias *et al.* [33] found gender bias among the network gatekeepers, where men get much greater attention compared to women. The authors also designed approaches to mitigate this gender bias. By analyzing the sharing behaviors of network gatekeepers, Orellana-Rodriguez *et al.* [37] proposed a set of guidelines to maximize engagement with the shared news items. Complementary to all these prior works, in our present work, we attempt to distinguish between the network gatekeepers who consume and share clickbait and traditional news stories.

## 2.5 Curiosity gap and psychological appeal of clickbaits

Media organizations have always struggled to bridge the gap between what the news producers tend to promote, and what the news readers are actually choosing to read. Some examples include journalists' penchant for public affairs while readers showing a reduced interest in the same. To overcome such problems, news producers often tend to use psychological methods to appeal the readers. The methods include creating *curiosity gap* [32] in the headline of the news article itself. Few recent works have attempted to understand such psychological appeals of clickbaits.

Specifically, researchers have examined how clickbaits employ two forms of forward referencing – *discourse deixis* and *cataphora* – to lure the readers to click on the article links [9]. Typically, this includes using pronouns to make reference to forthcoming parts in the discourse (often in the article text). For instance, the clickbait headline “This Twitter User Says She Was Suspended After Criticising Taylor Swift” is cataphoric because here ‘This’ refers to the name of the Twitter user, which is revealed only in the article body. In such headlines, there remain empty slots in the readers' mind that cannot be filled without going through the article in question [9]. These psychological tactics that clickbaits are known to engage, have become the crux of the arguments in favour of labeling clickbaits as misleading content or false news [15].

## 2.6 Automatic detection of clickbaits

There have been some recent attempts to detect and prevent clickbaits. Facebook attempted to remove clickbaits depending on the click-to-share ratio and the amount of time spent on different stories [39]. The browser plugin ‘Downworthy’ [26] detects clickbait headlines using a fixed set of common clickbaity phrases, and then converts them into meaningless garbage. Potthast *et al.* [40] attempted to detect clickbaits in Twitter by using common words occurring in clickbait tweets. Biyani *et al.* [7] proposed approaches to detect clickbaits using article informality. Anand *et al.* [3] used deep learning based techniques to detect clickbaits.

In our prior work [14], we compared clickbaits and traditional news headlines, and noticed that clickbait headlines use several language traits to attract users. For example, such headlines have more function words, more stopwords, more hyperbolic words, more internet slangs, and more frequent use of possessive case, as compared to the traditional headlines where the title contains specific proper nouns and the reporting is in third person [14]. Based on these observations, we developed a clickbait classifier where given a news article headline, the classifier would classify it as clickbait or non-clickbait. In our present study, we extend the classifier developed in [14] to separate clickbait and non-clickbait tweets.

However, the research questions we investigate in this work are complementary to the earlier work. For example, in [14], we identified linguistic characteristics that differentiate clickbait and



traditional news headlines. Whereas, in this paper, we explore complementary questions specific to tweets such as whether clickbait tweets contain several entities which might lead to their increased visibility, or whether the sentiment conveyed by the clickbait tweets differ from the non-clickbait tweets. Moreover, taking a very different direction compared to [14], we study the production and consumption patterns of clickbaits in Twitter, and bring out interesting insights.

### 3 DATASET GATHERED

As mentioned earlier, in this work, we attempt to analyze the usage as well as the users of clickbaits in social media. Towards that end, we gathered extensive longitudinal data from Twitter, covering a period of 8 months from February, 2016 to September, 2016. Throughout this 8 month period, using the Twitter Streaming API<sup>3</sup>, we collected all tweets posted by the Twitter handles of several (i) traditional news media organizations, and (ii) digital media outlets known to often deploy clickbaits.

As news media organizations, we considered the top three newspapers according to the Alexa ranking<sup>4</sup>: New York Times, Washington Post and India Times. Additionally, we considered one *online only* news media outlet – Huffington Post. Interestingly, these media organizations do not maintain a single Twitter account. Rather, alongside the primary account (i.e., @nytimes, @washingtonpost, @indiatimes, and @HuffPost), they also maintain several secondary accounts to tweet about stories related to specific news sections. For example, New York Times maintains more than 20 Twitter accounts (e.g., @nytpolitics, @nytnational, @nytimesworld, @nytopinion, @nytimesbusiness etc.). Similarly, Washington Post maintains more than 10 Twitter accounts (e.g., @postpolitics, @PostWorldNews, @PostOpinions, @PostSports etc.). In total, we collected tweets posted by the four primary accounts, and 38 secondary Twitter accounts of the media organizations in this category.

Regarding the media outlets promoting clickbaits, we considered the five outlets identified in our earlier work [14]: BuzzFeed, Upworthy, ViralNova, ScoopWhoop, and ViralStories. Additionally, we identified three more outlets: MensXP, 9GAG, and CountryLiving. We collected the tweets posted by their corresponding Twitter handles: @BuzzFeed, @Upworthy, @ViralNova, @ScoopWhoop, @allviralstories, @MensXP, @9GAG, and @CountryLiving. Similar to the traditional media organizations, many of these outlets also maintain multiple Twitter handles. For example, BuzzFeed maintains @BuzzFeedPol, @BuzzFeedNews, @BuzzFeedFashion etc. alongside @BuzzFeed. Similarly, Scoopwhoop maintains several secondary Twitter accounts like @ScoopwhoopONN, @scoopwhoopnews, @ScoopWhoopVideo etc. In total, we collected the tweets posted by 27 Twitter handles, which include the primary accounts as well as the secondary accounts of the media outlets in this category.

For both traditional media organizations and the outlets promoting clickbaits, in addition to the original tweets posted by them, we also collected all retweets (of these original tweets) made by their followers on Twitter. In total, we collected around 288K original tweets and 11.4M retweets posted during February, 2016 to September, 2016. Then, we attempted to separate the data into two categories: (i) a corpus of clickbait tweets, and (ii) another corpus of non-clickbait tweets.

In this work, we used a slightly modified version of the clickbait classifier developed in [14]. In [14], we identified several linguistic features which distinguish clickbait and traditional news headlines. These features include the length of the headline, the ratio between the number of stopwords to the number of content words, length of the longest separation between syntactically dependent words, the presence of cardinal numbers in the beginning of the headline, or the presence of unusual punctuation patterns and contracted word forms. Additionally, common clickbait phrases,

<sup>3</sup>[dev.twitter.com/streaming/overview](https://dev.twitter.com/streaming/overview)

<sup>4</sup>[alexa.com/topsites/category/News/Newspapers](https://www.alexa.com/topsites/category/News/Newspapers)

Twitter Handle	Example Clickbait Tweets
@BuzzFeedMusic	This theory about the new Radiohead album is driving fans crazy <a href="https://t.co/W26Glw5pEQ">https://t.co/W26Glw5pEQ</a> <a href="https://t.co/N9q5kIgsax">https://t.co/N9q5kIgsax</a>
@HuffingtonPost	39 breastfeeding portraits that celebrate nursing mamas <a href="https://t.co/D2XKXBeVmY">https://t.co/D2XKXBeVmY</a> <a href="https://t.co/PG48j7Dbuw">https://t.co/PG48j7Dbuw</a>
@BuzzFeed	Can You Guess Which One Of These People Is Holding A Vibrator? <a href="https://t.co/AhWF45yhzA">https://t.co/AhWF45yhzA</a>
@ScoopWhoop	One of these places is in #Switzerland. Can you guess which one it is?: <a href="https://t.co/7VRufhbaMp">https://t.co/7VRufhbaMp</a> <a href="https://t.co/nuh5n9xSnR">https://t.co/nuh5n9xSnR</a>
@Upworthy	These are 7 things they don't tell you about living with PCOS. <a href="https://t.co/CbwJSjtw4K">https://t.co/CbwJSjtw4K</a> <a href="https://t.co/GHtmKu5llw">https://t.co/GHtmKu5llw</a>
Twitter Handle	Example Non-clickbait Tweets
@HuffPostPol	Texas Lt. Gov. blames Black Lives Matter, social media for Dallas shooting <a href="https://t.co/wZRA0tId5e">https://t.co/wZRA0tId5e</a> <a href="https://t.co/7v5y3NzUAU">https://t.co/7v5y3NzUAU</a>
@NYTNational	Milwaukee joins a growing list of cities to experience police-related racial violence. <a href="https://t.co/ymPdXrVhNz">https://t.co/ymPdXrVhNz</a> <a href="https://t.co/CbZkfa2J4B">https://t.co/CbZkfa2J4B</a>
@nytimesworld	How Kurds backed by the Pentagon ended up fighting Syrian Arabs backed by the CIA in Syria <a href="https://t.co/on3XgfhLnl">https://t.co/on3XgfhLnl</a> <a href="https://t.co/JnkxTg4kgQ">https://t.co/JnkxTg4kgQ</a>
@BuzzFeedWorld	EgyptAir Flight Carrying 69 People Disappears From Radar <a href="https://t.co/XrRFYEIPCP">https://t.co/XrRFYEIPCP</a>
@nytpolitics	U.S. Strike in Yemen Kills Dozens in Qaeda Affiliate, Officials Say <a href="https://t.co/PXhABcCdlF">https://t.co/PXhABcCdlF</a>

Table 1. Few examples of clickbait and non-clickbait tweets in our dataset.

internet slangs and determiners, word N-grams, POS N-grams and Syntactic N-grams were also used as features for the SVM classifier distinguishing clickbaits and non-clickbaits.

In order to apply the classifier to classify tweets, we removed the hashtags and urls from the tweets, and only considered the tweet text for classification. We also excluded a particular feature used in the original classifier: length of the news headline (here, number of words in a tweet). In Twitter, all tweets are restricted to 140 characters, and therefore, we didn't find the length feature to be a major distinguisher between clickbait and non-clickbait tweets. We retained all other features used in the original classifier, and applied it to separate clickbait and non-clickbait tweets.

To evaluate the performance of the classifier, we asked three volunteers to manually label 200 randomly selected tweets from the dataset as either clickbait or non-clickbait. The inter-annotator agreement was 'substantial' with Fleiss'  $\kappa$  score [21] of 0.72. Taking the majority vote as ground truth, 88 tweets were identified as clickbait and rest of the tweets as non-clickbait. Then, we used the classifier, as described above, to classify the same tweets into clickbait and non-clickbait categories. By comparing the class predicted by the classifier, and the ground truth labels assigned by the human volunteers, we find the precision of the classification to be 0.872, recall 0.843, and F1-score to be 0.857. The accuracy of the classifier was 87.6% (against the random guessing accuracy of 50%).

As mentioned earlier, we applied this classifier on the tweets in our dataset (after removing hashtags and urls present in the tweets), and separated the data in two corpuses. The corpus of clickbait tweets consists of 126K original tweets, and more than 5M retweets posted by 1.08M unique users. On the other hand, the corpus of non-clickbait tweets contains 162K original tweets and 6.4M retweets posted by 1.29M unique users. Table 1 shows few example tweets from both clickbait and non-clickbait corpus. There are around 34K users who retweeted both clickbait and non-clickbait tweets. For our analysis, we do not consider these users and their retweets, and

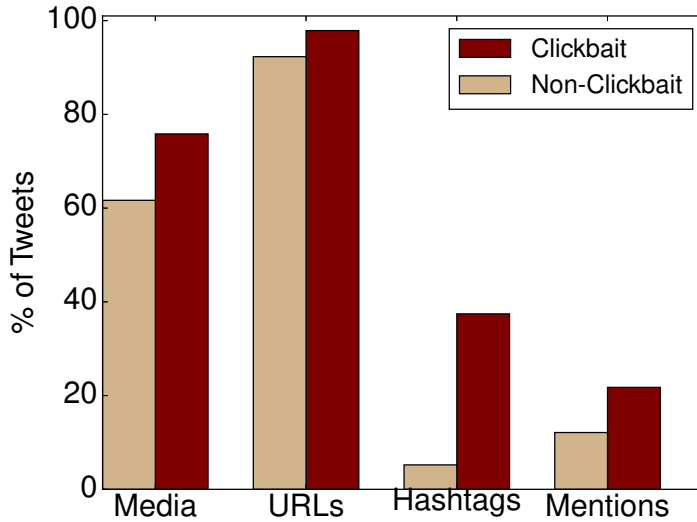


Fig. 1. Comparing the presence of different entities in both clickbait and non-clickbait tweets. We conducted the analysis over tweets posted everyday during our 8 months data collection period, and present the average values here. All the differences are found to be statistically significant with  $p < 0.05$  in Welch's T-tests.

only consider the users who exclusively participated in retweeting either clickbait or non-clickbait tweets in our dataset.

After gathering the tweets and retweets belonging to both clickbait and non-clickbait categories, in the subsequent sections, we attempt to answer the research questions we set out to investigate in this paper.

#### 4 RQ1. HOW ARE CLICKBAIT TWEETS DIFFERENT FROM NON-CLICKBAIT TWEETS?

In this section, we investigate how the clickbait and non-clickbait tweets differ from each other in terms of the tweet contents. Generally, we observe that clickbait tweets often mimic the article headlines, and both of them provide little information about the actual article contents. Here, the headlines and the tweets contain forward referencing cues to create a **curiosity gap** [32] to lure the readers so that they click on the article link to know more about the article. Whereas, non-clickbait tweets provide a summary of the content of the articles being referred to.

In the world of journalism, the process of the arousal of curiosity gap creates sensationalism. Just like clickbaits, any sensationalized content thrives on shock value. For this reason, tabloids in the past are known to use bold, attractive but sometimes misleading headlines. Such headlines contain something that the readers are already familiar with, and would want to know more [24]. In the subsequent subsections, we investigate the social media scenario – whether the clickbait tweets, along with the arousal of curiosity, employ additional techniques to reach a wide user base.

##### 4.1 Presence of different entities in the tweets

Apart from the main textual content, a tweet also contains other entities such as images, videos, URLs, hashtags and mentions of other users. Including media content (in the form of images and videos) in the tweets help in increasing its attractiveness and amplifying the curiosity among the



readers. Twitter provides an ubiquitous search box to search for any hashtag, and returns a list of tweets containing that particular hashtag. Therefore, including hashtags in the tweets help in reaching a wider audience on Twitter organically.

We start by comparing the presence of these different entities in both clickbait and non-clickbait tweets. Fig. 1 shows the percentage of tweets in these two categories with at least one of the corresponding entities present, where the differences between the two categories are statistically significant ( $p < 0.05$  in Welch's T-tests [46]). We can see in Fig. 1 that the percentage of clickbait tweets containing media contents, user mentions, URLs and hashtags are significantly higher than that of non-clickbait tweets.

While considering the average number of entities per tweet, we find that the clickbait tweets contain almost nine times the average number of hashtags (0.37) as compared to non-clickbait tweets (0.041). Similarly, clickbait tweets also contain a higher number of user-mention per tweet (0.427) compared to the non-clickbait tweets (0.305). Performing *Welch's T-tests* [46] over the entire corpus of clickbait and non-clickbait tweets confirm that the differences are *statistically significant* with  $p < 0.05$ .

Increased use of media elements (such as images) in clickbait tweets can be compared with the tabloids, which brought in a sudden increase in the number of visual elements in newspaper design. Similar to clickbaits, tabloids have also been known to edit and enhance photographs to suit the tone of the news, and reserve a greater space for colorful graphic advertisements [41].

Similarly, increased use of user mentions, as well as the use of hashtags, not only lead to a greater organic reach, but also make a personal connect with the audience. A similar attempt can be witnessed in tabloids as well, when a public content is presented in a personal way, be it the story of a victim, or the private matters in the life of an individual (e.g., affairs, relationships, disagreements etc). By doing so, tabloids ensure that the news has special relevance for the isolated reader [10].

## 4.2 Sentiment conveyed by the tweets

Now, we attempt to analyze the sentiments conveyed through both non-clickbait and clickbait tweets. To determine the sentiments of the tweets, we used the sentiment analysis tool developed by Rouvier *et al.* [44], which uses an ensemble of Convolutional Neural Networks (CNN) with embeddings trained on different units: lexical, part-of-speech, and sentiment embeddings. In the final fusion step, the hidden layers of the CNNs are concatenated and fed into another deep neural network for sentiment detection [44]. Given a tweet, it classifies the sentiment of the tweet into one of the three classes: positive, neutral, and negative.

To provide a benchmark for sentiment analysis of tweets, ACL SIGLEX ([siglex.org](http://siglex.org)) organizes *International Workshop on Semantic Evaluation (SemEval)*, which runs the competitive *Sentiment Analysis in Twitter* task every year [35]. The tool developed in [44] ranked 2nd at the SemEval task of 2016, achieving average F1 score of 63% over three sentiment categories. In this work, we used the tool developed by Rouvier *et al.* [44] primarily because of its superior performance, and also because the authors have kindly made the tool publicly available<sup>5</sup>. While analyzing the sentiment of clickbait and non-clickbait tweets, we first exclude the URLs associated with the tweets, and discard those tweets whose remaining lengths are less than three words.

We find that a higher fraction of non-clickbait tweets are associated with negative sentiment (26.12%) as compared to clickbait tweets (17.65%). Whereas, 40.75% clickbait tweets convey positive sentiments compared to only 15.44% non-clickbait tweets. This might be attributed to the fact that the news organizations tend to report on events which denote significant disruptions in the society, and often such news contain high negative sentiments. So, the proverb goes: *No news is good news*,

<sup>5</sup>available at [gitlab.lif.univ-mrs.fr/mickael.rouvier/SemEval2016](https://gitlab.lif.univ-mrs.fr/mickael.rouvier/SemEval2016)

and good news is no news. On the contrary, clickbait tweets convey more positive sentiment by covering mostly soft human interest stories to please their consumers.

Quite contrary to clickbaits, the offline media has been historically chided about the negativity and over sensationalization in the content of tabloids. In fact, Turner [54] acknowledged that the tabloid press prefers information over entertainment, or accuracy over sensation in their representations that exploit their audience. Often their focus is on staged family conflicts in an issue of public interest (e.g. a politician's private life). Interestingly, opposite to tabloids, clickbaits try to garner more attention by avoiding negative content.

### Summary

In summary, we find that clickbait tweets contain significantly higher fraction of images, hashtags and user mentions compared to non-clickbait tweets, to engage with a broader set of audience. The tweet texts in clickbaits often offer little detail about the article being referred to, which leads to an increase in the curiosity of the consumer. As a result, clickbait tweets contain more external article links than non-clickbait tweets. Finally, clickbait tweets convey more positive sentiments than non-clickbait ones, in contrast to the offline tabloids which thrived by propagating more negative sentiments.

## 5 RQ2. HOW DO CLICKBAIT PRODUCTION AND CONSUMPTION DIFFER FROM NON-CLICKBAITS?

After studying the tweet contents, the next question we investigate is the difference between the production and consumption of clickbait and non-clickbait tweets. We define the *producers* of clickbait tweets as those users who have originally posted the tweets, and *consumers* as the users who have retweeted a clickbait tweet. A similar definition is established for the producers and consumers of non-clickbait tweets. We identify retweets by investigating the 'retweeted\_status' field in the tweet metadata returned by Twitter API. If a particular tweet is a retweet of another tweet, then this field contains information about the original tweet that was retweeted. However, if some users copy and paste other tweets instead of retweeting them, then we can not track such duplicated tweets. We also acknowledge the limitation of using retweeting as a proxy for consumption. However, there is no direct way available to us to measure consumption at a large scale.

To answer RQ2, we first study the diurnal variation in the activities of the producers and the consumers of both clickbait and non-clickbait tweets.

### 5.1 Diurnal variations in tweeting and retweeting

We attempt to observe the diurnal variations in the activities of both clickbait and non-clickbait producers (mostly comprising different media sources), i.e., we want to see how active they remain at different times of the day. This investigation would tell us whether a specific category of tweet (e.g., clickbait) is mostly posted during leisure hours.

For both clickbaits and non-clickbaits, different Twitter accounts may be operated from different timezones. Hence, we convert the posting times of every tweet to the corresponding local time of the user posting them (by using the timezone information of the user as returned by Twitter API). Similarly, as the consumers of these tweets may be distributed worldwide, we also convert the retweet times to the users' respective local times. The normalization helps in a better comparison between the activities across timezones.

Fig. 2(a) shows the hourly variations in the production of both clickbait and non-clickbait tweets. Only except 9 AM, the fractions of clickbait and non-clickbait tweets being produced in all other hours are *significantly* different ( $p < 0.05$  in the corresponding Welch's T-tests). Similarly, Fig. 2(b)

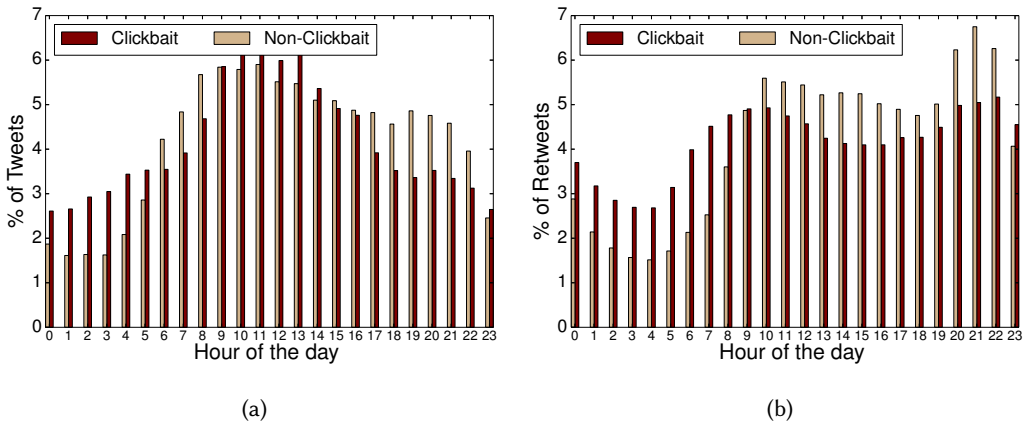


Fig. 2. Diurnal variations in (a) tweeting (i.e., production) and (b) retweeting (i.e., our proxy for consumption) of both clickbait and non-clickbait tweets posted everyday during our 8 months data collection period. Only except 9 AM, the differences in average values for all other hours are found to be statistically significant with  $p < 0.05$  in Welch's T-tests.

shows the hourly distributions of retweeting activity in both clickbait and non-clickbait tweets. Even in case of retweets, we observe significant difference in hourly activities of clickbait and non-clickbait consumers.

We see some common features between the overall patterns in both production and consumption activity in Fig. 2(a), and in Fig. 2(b). For instance, both production and consumption is relatively less during the night (10 PM to 5 AM) for both clickbaits and non-clickbaits compared to their day activities. However, we also notice some interesting differences in the production and consumption activities of the two categories. For instance, clickbait tweets tend to be posted more during peak hours of the day (from 10 AM to 2 PM) as well as on night (from 11 PM to 5 AM). Whereas, consumption of clickbait tweets is higher in the odd hours of the night and early morning (between 11 PM and 8 AM), but for the rest of the day, users tend to consume more non-clickbait tweets.

## 5.2 Longevity of individual tweets

Next, we attempt to measure how long the consumers remain engaged with individual clickbait and non-clickbait tweets. Towards that end, we chronologically order the retweets for a tweet, and we define the *longevity of a tweet* as the difference between the time when the original tweet was posted and the time of the 95th percentile retweet of the original tweet (in our dataset). Fig. 3(a) and Fig. 3(b) show the longevity of different tweets in both clickbait and non-clickbait categories. We can see clear distinction between the engagement levels in clickbait and non-clickbait tweets. We find that clickbait tweets have higher user engagement levels as compared to that of non-clickbait tweets. The spikes in the longevity of clickbait tweets show that higher number of clickbait tweets received retweets beyond 50 hours from the posting of the original tweet.

Such difference in engagements possibly arise because the non-clickbait tweets tend to focus on some temporal events (e.g., political rallies, sports matches, entertainment events etc.), and hence the user attention decays rapidly with time as the relevance of the corresponding event fades away. On the other hand, there is less time binding associated with clickbait tweets, and therefore, they tend to attract user engagements beyond any specific time threshold.

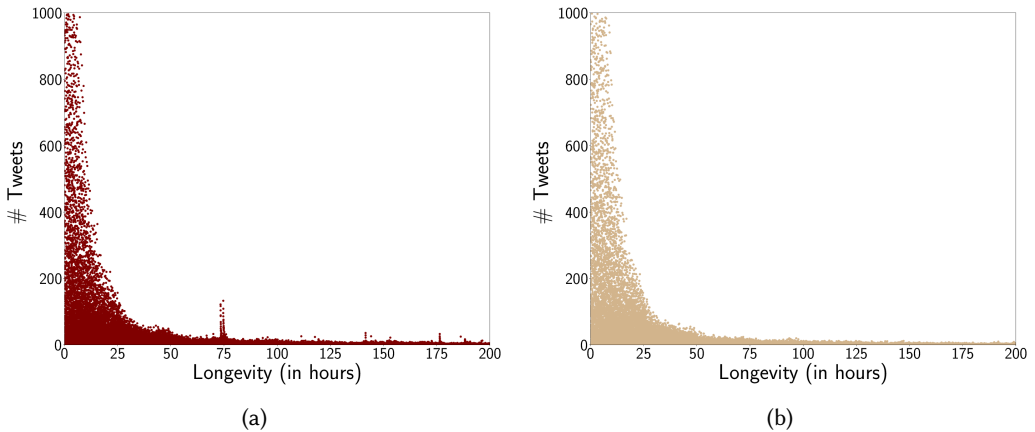


Fig. 3. Scatter plot of longevity values (in hours) for different tweets in (a) Clickbait, and (b) Non-clickbait category.

### 5.3 Length of the retweet cascade

One of the main characteristics of social networks like Twitter is that it allows its users to post and *reshare* content they think would be useful for their followers. In some cases, an original tweet gets re-(and-re-re)tweeted multiple times. For example, a user first posts a tweet and then one of her followers shares the tweet with her set of followers, and some of these followers retweet to their respective sets of followers, and so on. In this process, a *retweet cascade* [16] gets formed, thereby potentially reaching a large audience. Information cascades have been studied in multiple different settings such as *viral marketing* [30] or *email-chains* [31], and it has been noted that the larger the cascades are, higher reach they have in terms of audience size.

After observing that clickbait tweets tend to get user attention for longer duration, a related question to ask is whether the length of the retweet cascades formed in clickbait and non-clickbait tweets are similar. We calculate the cascade lengths of tweets belonging to both clickbait and non-clickbait categories, and find that most of the tweets in both categories have cascade length 1, denoting the retweets made within the direct followers. This finding corroborates with earlier studies which observed that large cascades are rare [30, 31]. However, on considering the 90-th percentile cascade lengths, we see that the clickbait tweets (with the value of 6.47) could penetrate the users to a deeper extent, and hence have a higher reach than the non-clickbait tweets (with the 90-th percentile cascade length being 5.23).

#### Summary

By conducting the above analyses, we see that the clickbait tweets retain their popularity among their user base much longer as compared to non-clickbait tweets. The followers of users retweeting clickbait tweets take active notice and subsequently participate in its propagation. This can largely be attributed to the fact that clickbait tweets are generally associated with topics which are not time-bound, and they seem more attractive to the consumers by virtue of the arousal of curiosity gap, as described earlier.

## 6 RQ3. WHO ARE THE CONSUMERS OF CLICKBAIT AND NON-CLICKBAIT TWEETS?

As mentioned in Section 2, prior readership surveys have found that the traditional broadsheet newspapers mostly cater to the affluent and well educated audience. Whereas, the majority of

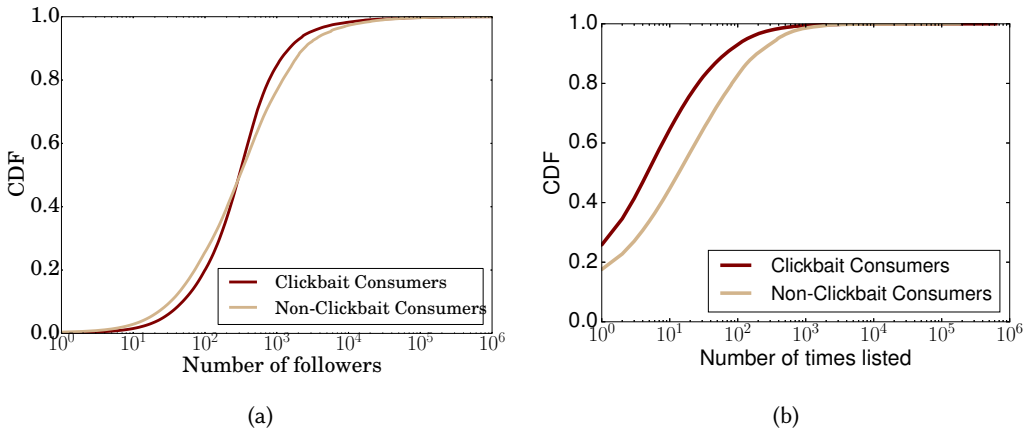


Fig. 4. For both clickbait and non-clickbait consumers, CDF of (a) number of their followers, and (b) number of times they are listed.

the tabloid readers are found to be in the lower middle and working class. Following that line, in this section, we try to understand the distinction between the users who consume clickbait and non-clickbait tweets. We compare and contrast their popularity and reputation on social media (which can act as a proxy to their class information), and their demographics. By analyzing the popularity and reputation, we hope to conclude whether a type of tweet is consumed by more popular users (users having high follower count) or by less popular users. Also, by studying the demographics of consumers such as their gender and age, we expect to get an account of what section of the society follows clickbait or non-clickbait tweets.

### 6.1 Popularity and reputation of consumers

Next, we turn our focus towards understanding how popular the consumers of clickbait tweets are vis-a-vis the consumers of non-clickbait tweets. Depending on the number of followers they have, we group the Twitter users into three different categories:

- (i) Less popular users (having less than 1,000 followers),
- (ii) Popular users (having 1,000 to 10,000 followers), and
- (iii) Very popular users (having more than 10,000 followers).

Fig. 4(a) shows the cumulative distribution function (CDF) of the number of followers for both clickbait and non-clickbait consumers. We can see in Fig. 4(a) that among the less popular range, on average, clickbait consumers have more followers than non-clickbait consumers. In other words, among the less popular users, clickbait consumers tend to be followed more than non-clickbait consumers. However, this trend reverses among popular users, where non-clickbait consumers are more followed than clickbait consumers. Finally, we don't find much difference in number of followers of the very popular users for both clickbait and non-clickbait tweets.

In addition to comparing the number of followers, we also look at the number of lists in which the consumers are enlisted. *Lists* is a feature in Twitter, using which a user can create a list of other users who tweet on a particular topic. For example, a user may create a list 'Politics' to include @realDonaldTrump, @BarackObama, @whitehouse, or @HillaryClinton. Similarly, another list 'Music' may include @britneyspears, @ladygaga, or @rihanna. The Twitter users present on these lists can be thought as the *experts* on that topic. More a particular user is enlisted in different lists, she can be regarded as more reputed in the Twitter community. There have been prior research works which have used these lists to build expert search and recommendation systems [25, 48].

In our context, we use this list feature to compare the reputation of both clickbait and non-clickbait consumers. More specifically, we compare the number of times different consumers are enlisted in different lists. As can be seen in the CDF plot in Fig. 4(b), non-clickbait consumers are listed much more on average than the clickbait consumers. Therefore, we can conclude that non-clickbait consumers are more reputed in the community than their clickbait counterparts.

## 6.2 What the profile descriptions reveal

Next, we analyze the self-identified Twitter biographies of the consumers. We perform an LIWC [53] analysis on the words used by the clickbait and non-clickbait consumers in their profile bios. We notice a stark difference in the use of several linguistic categories by clickbait consumers and their non-clickbait counterparts (the differences are found to be *statistically significant* in Welch's T-tests [46]). For example, clickbait consumers tend to use swear words 130.77% more often than non-clickbait consumers (0.3 vs 0.13,  $p < 0.05$ ). Additionally, the use of personal pronouns by clickbait consumers is 30.04% more than non-clickbait consumers (6.71 vs 5.16,  $p < 0.05$ ). We also note a significant jump in the use of words with negative tones by clickbait consumers. The use of anxiety words like nervous, afraid and tense is 95.24% more by clickbait consumers than by their non-clickbait counterparts (0.41 vs 0.21,  $p < 0.05$ ). They are also 63.16% more frequent in using words like grief, cry and sad (0.31 vs 0.19,  $p < 0.05$ ). Clickbait consumers use 80% more filler words like 'you know' and 'I mean' (0.18 vs 0.1,  $p < 0.05$ ).

Non-clickbait consumers, on the other hand, are 27% more likely to use family related words (1.07 vs 0.84,  $p < 0.05$ ), and are 56% more frequent in using words related to work, achievements and their occupational roles (e.g., *lawyer*, *scientist*, *reporter*) (5.93 vs 3.8,  $p < 0.05$ ). Moreover, their interests seem to be in the areas of business and economic affairs, as reflected by their 65% higher use of words like audit, cash and owe (pertaining to money) (1.09 vs 0.66,  $p < 0.05$ ). We also notice words like *girl* and *princess* appearing more among the bio of clickbait consumers; while the word *father* appearing more in its non-clickbait counterpart. This indicates a tilting of women towards clickbait articles and more male prominence among the non-clickbait consumers. We now look at the demographic distribution of the consumers to more thoroughly investigate this tilt.

## 6.3 Consumer demographics

To compare the demographics of the clickbait and non-clickbait consumers, we consider their gender, and age. However, inferring the demographics of Twitter users at scale is challenging. There have been some approaches for inferring the gender of a Twitter user from the username [8], or the age from Twitter profile description (by finding textual patterns like '21 yr old', 'born in 1989') [51]. However, due to their reliance on standard usernames and profile descriptions, these approaches fail to infer the demographics for a large number of users [13]. Following the approach taken by some of the earlier attempts in working with user demographics [2, 13], we use *Face++ API* [28], a face recognition platform based on deep learning [58], to extract the gender and age from the recognized faces in the profile images of the Twitter users. Some profile images may not have any recognizable face, while some other images may have more than one faces. Discarding these images (and the image URLs which were unavailable at the time of the inference), we end up getting the demographic information for around 208K clickbait consumers and 226K non-clickbait consumers. Face++ returns 'Male' or 'Female' as the gender, and a numerical value as the age of a user.

To evaluate the performance of Face++, we first collected a set of users for whom we can infer the gender and age from their screen names and profile descriptions. Among the users for whom we got the inferred demographics from Face++, we could successfully get the gender from the screen name (using the method developed by Blevins *et al.* [8]), and age from the profile description (using the method proposed by Sloan *et al.* [51]) for 3,208 users. Then, we computed the accuracy



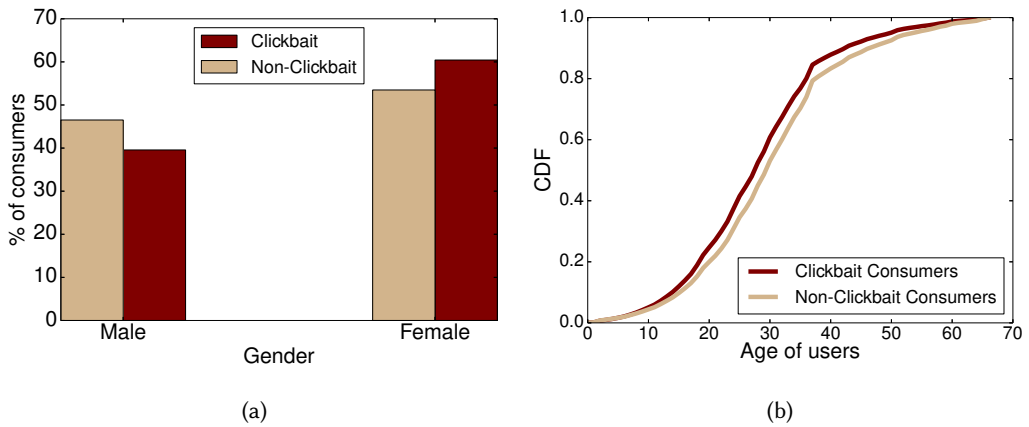


Fig. 5. **Demographics of clickbait and non-clickbait consumers: (a) their gender distribution, and (b) CDF of their age.**

of the demographic inference by Face++, by considering the data obtained for these 3, 208 users as ground truth. We found the gender inference accuracy of Face++ to be 84%; whereas, the average error in the age inferred by Face++ was 5.2 years. We note that the accuracy results are similar to the ones reported in earlier evaluations [2, 13].

Using the demographic information returned by Face++ for both clickbait and non-clickbait consumers, we compute their overall demographics considering their gender and age. Fig. 5(a) shows the gender distribution among both the clickbait and non-clickbait consumers. Fig. 5(b) shows the CDF of ages of the consumers of both categories.

In a recent work, Chakraborty *et al.* [13] reported that among the Twitter users, there are more women than men (53.1% vs. 46.9). Therefore, we can see in Fig. 5(a) that for both clickbait and non-clickbait consumers, there are more women than men. However, by comparing the fraction of women among the clickbait and non-clickbait consumers, we see that there are higher fraction of women among clickbait consumers than non-clickbait consumers. Whereas, it is the opposite when comparing the fraction of men among the consumers.

Surprisingly, this observation is contrary to the gender distribution observed among tabloid readers, where the majority of readers are men [29]. Journalism researchers have argued that even though most of the tabloids provide woman's pages reporting on fashion and woman's health issues in addition to the celebrity gossip pages, sexualised representation of women in the overall news discourse attracts way more male readers to tabloids [29].

Regarding the age of the consumers, Fig. 5(b) shows that the clickbait consumers tend to be younger than the non-clickbait consumers. Similar to clickbaits, a large fraction of tabloid readers are aged below 35 years, compared to much less fraction of readers aged above 65 years [29].

## Summary

By conducting the analyses in this section, we find that the clickbait tweets are more popular among women, and also among relatively younger Twitter users. On the other hand, non-clickbait consumers have a higher proportion of men and older people. Additionally, they are more reputed, and have relatively higher follower base. Profile descriptions of both type of consumers reveal significant differences in the use of words from different linguistic categories.

## 7 RQ4. HOW DO THE CLICKBAIT AND NON-CLICKBAIT CONSUMERS DIFFER AS A GROUP?

Criticisms around tabloidization are mostly grounded in Habermas' notion of the *public sphere* [27], where public opinion can be formed via *rational-critical debates* between private individuals. Although news media is considered to be the enabler of such communications, critiques of tabloids have argued that tabloids fail those standards to enable debates in the public sphere [29].

In this work, we investigate a related question in the context of social media – among clickbaits or non-clickbaits, which is a better enabler of communication between different groups of people in the public sphere (i.e., the social media platform). To answer this question, we study the reciprocity of the follower graph, retweet graph, and mention graph constructed from the activities of both clickbait and non-clickbait consumers, where reciprocity determines the extent of mutual engagement between the consumers. We also analyzed the density of the graphs, but the graphs turn out to be very sparse for both type of consumers.

### 7.1 Properties of the follower graph

We can visualize the full Twitter network as a directed graph where every user has a link to her follower. Reciprocity in this graph is the fraction of user pairs having links both ways (i.e., they have bidirectional following relation). The follower graph of clickbait and non-clickbait consumers are two induced subgraphs from the large Twitter follower network. We calculated the reciprocity in both subgraphs. However, we did not record significant difference in the results obtained for the two kinds of consumers. For the clickbait consumers, the reciprocity value is 0.00187, while it is 0.00186 for the non-clickbait consumers. The low reciprocity suggests that forging mutual following is not very common for both clickbait and non-clickbait consumers.

### 7.2 Properties of the retweet graph

We next analyze the retweet graph of clickbait and non-clickbait tweets. Two users in a retweet graph are connected by bidirectional edge when both have retweeted some tweets posted by the other. We find that the reciprocity in case of clickbait consumers ( $9.74e^{-4}$ ) is almost 1.5 times that of the non-clickbait consumers ( $6.04e^{-4}$ ). The results tell us that the clickbait consumers tend to mutually engage more in terms of retweeting each other than non-clickbait consumers.

### 7.3 Properties of the mention graph

Similar to the retweet graph, in the mention graph, bidirectional edge is created when two users mention each other in their tweets. We next analyze the mention graph of clickbait and non-clickbait consumers. We find that the reciprocity for clickbait consumers ( $1.63e^{-3}$ ) is almost 2 times that of the non-clickbait consumers ( $8.17e^{-4}$ ). Therefore, similar to the retweet activity, clickbait consumers also tend to mention each other way more than their non-clickbait counterparts.

### 7.4 Retweet frequency of the users

We now analyze the retweeting activities of both clickbait and non-clickbait consumers. To this end, we first rank the consumers based on the number of retweets they made, and then plot the ranks against the corresponding retweet count in Fig. 6. We observe in Fig. 6 that the retweeting activities of both clickbait and non-clickbait consumers exhibit *heavy tailed* distributions where a small number of users have made most of the retweets.

However, the distribution for clickbait consumers are relatively less skewed than that of non-clickbait consumers. This is interesting because when we compute the average number of retweets made, we find similar values for both clickbait (5.257), and non-clickbait consumers (5.165). Such

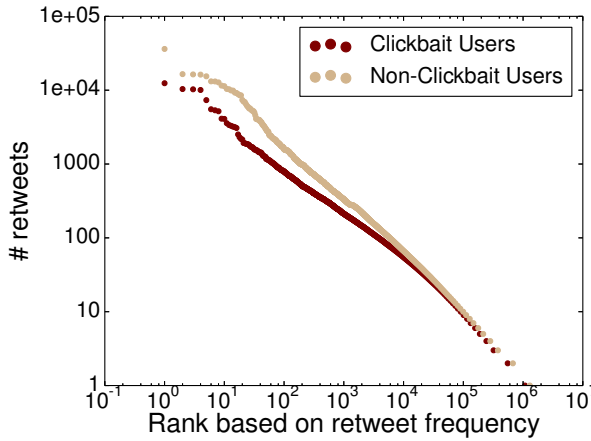


Fig. 6. Retweet frequencies of clickbait and non-clickbait consumers.

higher skew in non-clickbait consumers signifies the potential presence of a core group of consumers who are more active in retweeting non-clickbait tweets. Members of such core groups can play the roles of *opinion leaders* in the public sphere [42], and initiate discussions in their communities around the non-clickbait stories. On the other hand, clickbait consumers have relatively more uniform retweeting activities implying less reliance on opinion leaders and more democratization of the discussions around clickbait stories.

### Summary

Doing the above analyses, we find that the level of mutual engagement between clickbait consumers is more than that of non-clickbait consumers. Also, we find that clickbait consumers follow a much uniform pattern while retweeting as compared to the non-clickbait consumers. For long, there have been many criticisms of Habermas' notion of public sphere. For example, Fraser argued that Habermas' conception is elitist, and therefore, members of subordinated groups require alternative arenas for *public discourse* to articulate their interests [23]. The idea of *alternative public spheres* has been extensively used to justify the effectiveness of tabloids in providing such arena [29, 38]. Likewise, the observations in this section highlight that in social media, clickbaits successfully provide such alternative public sphere for users drifting away from traditional news.

## 8 CONCLUDING DISCUSSION

In this work, we analyzed the production and consumption of clickbaits in social media. We posed four related research questions which we answered using several analyses. We observed that there is a clear distinction in the way clickbait tweets and their consumers differ from that of non-clickbaits.

Our investigation reveals several interesting insights. Clickbait tweets include more entities such as images, hashtags, and user mentions which help in capturing the attention of the consumers. It is also noted that a higher percentage of clickbait tweets convey positive sentiments as compared to non-clickbait tweets. Clickbait tweets tend to have a wider and deeper reach in its consumer base as compared to non-clickbait tweets.

Additionally, we made interesting observations regarding the clickbait consumers. For example, clickbait tweets are consumed by more women compared to their fraction among the consumers of non-clickbait tweets. Also, the clickbait consumers are younger than non-clickbait consumers. The results also point out that the non-clickbait consumers are more reputed in the community,

and the linguistic composition of their profile descriptions differ significantly from the clickbait consumers. It can also be concluded the clickbait consumers have more mutual engagement among each other, and they retweet more uniformly than non-clickbait consumers.

The conclusion that the clickbait tweets are more popular among a section of the society which is clearly distinct from the consumers of regular news, again corroborates the general notion that the mainstream journalism is more for a certain class of the society (e.g., affluent and higher-educated). For a long time in history, the narrative in newspapers was also controlled by that class (including politicians, filmstars, sportsmen, large business owners, administrators and general experts). In this context, the tabloids emerged as a useful tool for raising the societal awareness among the *subalterns* [5]. We can hypothesize that just like tabloids, clickbaits also cater to the section of the society which shows little interest in following mainstream journalism.

In fact, it is noted that the criticism of tabloid journalism and tabloid form in general is more often made using traditional criteria of political power (voting, participation in formal political activities etc.), rather than the criteria of cultural recognition (representation, participation in alternate political forms etc.) [38]. Similarly, prior works on clickbaits report only the negative aspects of it, by turning a blind eye to the vast majority of users who are embracing clickbaits. We can extend the earlier argument to point out that only criticizing clickbaits and trying to get rid of them altogether may not be desirable. Instead of blanket prevention of clickbaits, we believe that a four-pronged approach should be taken to tackle the prevalence of clickbaits.

First, an alternate solution may be to indicate the *newsworthiness* of an article to the intended readers. It can be integrated in form of the browser extensions developed in [14, 26]. When a user hover the mouse over an article link, she can be shown a short summary of the article being referred to, along with the possible news value of the article. Such an option would bring in more transparency to the users, and they can decide whether to pursue a particular article or not.

Second, if certain groups of users indeed pursue clickbaits more, dedicated clickbait recommender systems can be developed to deliver more values to them. In such systems, relatively better quality articles may be given higher importance, which can potentially weed out lower quality information and encourage clickbait producers to come up with attractive yet informative article content.

Third, clicks are important in today's advertisement driven revenue model. Hence, motivated by the success of clickbait headlines, many traditional media organizations have also started experimenting with *attractive* headlines to catch readers' attention. These organizations are deploying A/B tests by exposing different headlines to different readers to identify the one capturing the readers' attention most<sup>6</sup>. It is to be seen whether such strategies can bring some of the clickbait consumers to the traditional news media landscape. Additionally, to have more engagements with social media audience, traditional media organizations can adopt some strategies from clickbait media's playbook. They can use more pictures and hashtags in their tweets. They can also utilize the user mentions and replies to establish direct connections with their intended audience. Such outreach can help in getting more clicks, and also reaching the users who feel left-out in the traditional news discourse.

Finally, in certain scenarios, clickbait prevention may be required. For example, there has been a recent development central to any discussion around news stories in social media – 'Fake News' [55]. Debates are going on regarding the spread of many false and misleading articles in social media, and its role in shaping public opinion during election periods [1]. To attract readers, such fake news articles often deploy catchy sensational headlines. Similarly, articles propagating extreme opinion also deploy clickbait techniques to get more views. In such situations, going beyond detecting

---

<sup>6</sup>[nytimes.com/2016/06/13/insider/which-headlines-attract-most-readers.html](https://www.nytimes.com/2016/06/13/insider/which-headlines-attract-most-readers.html)

clickbaits, we may need to distinguish between such ‘bad’ vs potentially ‘good’ clickbaits, and apply prevention techniques (e.g., methods proposed in [14]) to stop widening their reach.

The solutions proposed above are only a few possible options to tackle the lowering of news value with the prevalence of clickbaits. Further research is needed to explore other alternatives. The study of clickbaits is still in its nascent stage, and there are a lot of open questions for future pursuits. For example, conducting in person interviews with the most active consumers of both clickbaits and non-clickbaits will help in understanding the consumer’s perspectives on the usefulness of clickbaits as well as traditional news.

Additionally, there is a need to compare more fine-grained user characteristics in future studies. For example, different business organizations maintain Twitter accounts to communicate with their consumers [47]. Such organizational accounts form a substantial group on Twitter. It might be interesting to compare the presence of organizational accounts among clickbait and non-clickbait consumers. The non-clickbait consumers may involve more organizational accounts, while the clickbait consumers may include more individual users. Similar to organizational accounts, in a recent study, Varol *et al.* [56] claimed that as much as 15% of Twitter accounts may be bots rather than real people. It is also worth investigating, whether certain media outlets have many bots among their consumers, which in turn may help in boosting their popularity in the eyes of the real consumers.

In conclusion, before making any decision on the fate of clickbaits in the future, it is of paramount importance to get the holistic picture and assess its social implications. In this work, we have made the first attempt towards that direction, and we hope that it will trigger discussions similar to what happened around the tabloidization in the offline media.

## REFERENCES

- [1] Hunt Allcott and Matthew Gentzkow. 2017. *Social media and fake news in the 2016 election*. Technical Report. National Bureau of Economic Research.
- [2] Jisun An and Ingmar Weber. 2016. # greysanatomy vs.# yankees: Demographics and Hashtag Use on Twitter. In *AAAI ICWSM*.
- [3] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. 2017. We used Neural Networks to Detect Clickbaits: You won’t believe what happened Next!. In *ECIR*.
- [4] Jeremy Barr. 2016. What Website? New Social Media-Only Brand Obsessee Hopes to Appeal to Teens. [adageindia.in/media/what-website-new-social-media-only-brand-obsessee-hopes-to-appeal-to-teens/articleshow/51318961.cms](http://adageindia.in/media/what-website-new-social-media-only-brand-obsessee-hopes-to-appeal-to-teens/articleshow/51318961.cms). (March 2016).
- [5] Matthew A Baum and Angela S Jamison. 2006. The Oprah effect: How soft news helps inattentive citizens vote consistently. *Journal of Politics* 68, 4 (2006).
- [6] S. Elizabeth Bird. 2009. Tabloidization: What is it, and Does it Really Matter?. In *The changing faces of journalism*, Barbie Zelizer (Ed.). Routledge.
- [7] Prakhar Biyani, Kostas Tsioutsoulouklis, and John Blackmer. 2016. 8 amazing secrets for getting more clicks: detecting clickbaits in news streams using article informality. In *AAAI*.
- [8] Cameron Blevins and Lincoln Mullen. 2015. Jane, John... Leslie? a historical method for algorithmic gender prediction. *Digital Humanities Quarterly* 9, 3 (2015).
- [9] Jonas Nygaard Blom and Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76 (2015).
- [10] Sally Brooke Cameron. 2010. *The Journal of Modern Periodical Studies* 1 (2010).
- [11] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2015. Can trending news stories create coverage bias? on the impact of high content churn in online news media. In *Computation and Journalism Symposium*.
- [12] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2016. Dissemination Biases of Social Media Channels: On the Topical Coverage of Socially Shared News.. In *AAAI ICWSM*.
- [13] Abhijnan Chakraborty, Johnatanan Messias, Fabricio Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2017. Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. In *AAAI ICWSM*.

- [14] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media. In *ACM/IEEE ASONAM*.
- [15] Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading Online Content: Recognizing Clickbait as False News. In *ACM MDD*.
- [16] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted?. In *ACM WWW*.
- [17] Michael X Delli Carpini and Bruce A Williams. 2001. Let us infotain you: Politics in the new media age. (2001).
- [18] Nicholas Diakopoulos and Arkaitz Zubiaga. 2014. Newsworthiness and Network Gatekeeping on Twitter: The Role of Social Deviance. In *AAAI ICWSM*.
- [19] Jeffrey Dvorkin. 2015. Column: Why click-bait will be the death of journalism. [pbs.org/newshour/making-sense/what-you-dont-know-about-click-bait-journalism-could-kill-you](http://pbs.org/newshour/making-sense/what-you-dont-know-about-click-bait-journalism-could-kill-you). (2015).
- [20] Martin Eide and Graham Knight. 1999. Public/private service: Service journalism and the problems of everyday life. *European Journal of Communication* 14, 4 (1999).
- [21] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971).
- [22] Ben Frampton. 2015. Clickbait: The changing face of online journalism. [bbc.com/news/uk-wales-34213693](http://bbc.com/news/uk-wales-34213693). (2015).
- [23] Nancy Fraser. 1990. Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social text* 25/26 (1990).
- [24] Herbert J Gans and B Zeilizer. 2009. Can popularization help the news media. *The Changing Faces of Journalism. Tabloidization, Technology, and Truthiness*. (2009).
- [25] Saptarshi Ghosh, Naveen Sharma, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. 2012. Cognos: crowdsourcing search for topic experts in microblogs. In *ACM SIGIR*.
- [26] Alison Gianotto. 2016. Downworthy: A browser plugin to turn hyperbolic viral headlines into what they really mean. [downworthy.snipe.net](http://downworthy.snipe.net). (2016).
- [27] Jürgen Habermas. 1991. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.
- [28] Megvii Inc. 2013. Face++ Research Toolkit. [www.faceplusplus.com](http://www.faceplusplus.com). (Dec 2013).
- [29] Sofia Johansson. 2007. *Reading tabloids: Tabloid newspapers and their readers*. Södertörns högskola.
- [30] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. 2007. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)* 1, 1 (2007).
- [31] David Liben-Nowell and Jon Kleinberg. 2008. Tracing information flow on a global scale using Internet chain-letter data. *PNAS* 105, 12 (2008).
- [32] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994).
- [33] J Nathan Matias, Sarah Szalavitz, and Ethan Zuckerman. 2017. FollowBias: Supporting Behavior Change toward Gender Equality by Networked Gatekeepers on Social Media. In *ACM CSCW*.
- [34] Amy Mitchell, Jeffrey Gottfried, Jocelyn Kiley, and Katerina Eva Matsa. 2014. Social Media, Political News and Ideology | Pew Research Center. [journalism.org/2014/10/21/section-2-social-media-political-news-and-ideology/](http://journalism.org/2014/10/21/section-2-social-media-political-news-and-ideology/). (Oct 2014).
- [35] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. SemEval-2016 Task 4: Sentiment Analysis in Twitter. In *SemEval @ NAACL-HLT*.
- [36] Rupert Neate. 2014. BuzzFeed valued at more than three times the Washington Post. [theguardian.com/media/2014/aug/11/buzzfeed-valued-at-three-times-washington-post](http://theguardian.com/media/2014/aug/11/buzzfeed-valued-at-three-times-washington-post). (2014).
- [37] Claudia Orellana-Rodriguez, Derek Greene, and Mark T Keane. 2016. Spreading the news: how can journalists gain more engagement for their tweets?. In *ACM WebScience*.
- [38] Henrik Ornebring and Anna Maria Jonsson. 2004. Tabloid journalism and the public sphere: A historical perspective on tabloid journalism. *Journalism Studies* 5, 3 (2004).
- [39] Alex Peysakhovich and Kristin Hendrix. 2016. News Feed FYI: Further Reducing Clickbait in Feed. [newsroom.fb.com/news/2016/08/news-feed-fyi-further-reducing-clickbait-in-feed/](http://newsroom.fb.com/news/2016/08/news-feed-fyi-further-reducing-clickbait-in-feed/). (2016).
- [40] Martin Potthast, Sebastian Köpse, Benno Stein, and Matthias Hagen. 2016. Clickbait Detection. In *ECIR*.
- [41] Divya Jyoti Randev. 2016. The Nature of Tabloidized Content in Newspapers: An Overview. (2016).
- [42] Marsha L Richins and Teri Root-Shaffer. 1988. The role of evolvment and opinion leadership in consumer word-of-mouth: An implicit model made explicit. *NA-Advances in Consumer Research* (1988).
- [43] Dick Rooney. 2000. Thirty years of competition in the British tabloid press. *Tabloid Tales: Global Debates over Media Standards – New York and Oxford* (2000).
- [44] Mickael Rouvier and Benoit Favre. 2016. SENSEI-LIF at SemEval-2016 Task 4: Polarity embedding fusion for robust sentiment analysis. In *NAACL 2016*.
- [45] David Rowe. 2011. Obituary for the newspaper? Tracking the tabloid. *Journalism* (2011).



- [46] Graeme D Ruxton. 2006. The unequal variance T-test is an underused alternative to Student's T-test and the Mann-Whitney U-test. *Behavioral Ecology* 17, 4 (2006).
- [47] Adam J Saffer, Erich J Sommerfeldt, and Maureen Taylor. 2013. The effects of organizational Twitter interactivity on organization–public relationships. *Public Relations Review* 39, 3 (2013).
- [48] Naveen Kumar Sharma, Saptarshi Ghosh, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. 2012. Inferring who-is-who in the Twitter social network. *ACM SIGCOMM Computer Communication Review* 42, 4 (2012).
- [49] Pamela J Shoemaker, Tim P Vos, and Stephen D Reese. 2009. Journalists as gatekeepers. *The handbook of journalism studies* 73 (2009).
- [50] Morten Skovsgaard. 2014. A tabloid mind? Professional values and organizational pressures as explanations of tabloid journalism. *Media, Culture & Society* 36, 2 (2014).
- [51] Luke Sloan, Jeffrey Morgan, Pete Burnap, and Matthew Williams. 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS one* 10, 3 (2015).
- [52] Mitchell Stephens. 2007. *A history of news*.
- [53] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010).
- [54] Graeme Turner. 1999. Tabloidization, journalism and the possibility of critique. *International Journal of Cultural Studies* 2, 1 (1999).
- [55] Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2017. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *new media & society* (2017).
- [56] Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107* (2017).
- [57] Kevin Williams. 2003. *Understanding media theory*.
- [58] Qi Yin, Zhimin Cao, Yuning Jiang, and Haoqiang Fan. 2015. Learning Deep Face Representation. (December 2015). US Patent 20,150,347,820.

Received April 2017; revised July 2017; accepted November 2017